Point- and Volume-based Multi-Object Acquisition in VR

Zhiqing Wu[1,2], Difeng Yu[1,3], and Jorge Goncalves[1]

[1] University of Melbourne, Australia
[2] Hong Kong University of Science and Technology (Guangzhou), China
[3] University of Copenhagen, Denmark

Abstract. Multi-object acquisition is indispensable for many VR applications. Commonly, users select a group of objects of interest to perform further transformation or analysis. In this paper, we present three multi-object selection techniques that were derived based on a two-dimensional design space. The primary design dimension concerns whether a technique acquires targets through point-based methods (selecting one object at a time) or volume-based methods (selecting a set of objects within a selection volume). The secondary design dimension examines the mechanisms of selection and deselection (cancel the selection of unwanted objects). We compared these techniques through a user study, emphasizing on scenarios with more randomly distributed objects. We discovered, for example, that the point-based technique was more efficient and robust than the volume-based techniques in environments where the targets did not follow a specific layout. We also found that users applied the deselection mechanism mostly for error correction. We provide an in-depth discussion of our findings and further distill design implications for future applications that leverage multi-object acquisition techniques in VR.

Keywords: Multiple targets · Object selection · Virtual Reality.

1 Introduction

The rapid development of virtual reality (VR) systems shows great potential for its use in practice, such as in industrial design, data exploration, and professional training [7,17,18]. Many of these application scenarios require the acquisition of multiple targets altogether. For example, as shown in Figure 1, a user wants to pick up all the bottles in a virtual chemistry lab or acquire data points with certain features in a 3D data cloud for further analysis.

To satisfy this particular need, many off-the-shelf VR applications, including Google Blocks, Tilt Brush, Tvori, Microsoft Maquette, and Gravity Sketch, have released a diverse range of multi-object acquisition techniques [25]. Some techniques require users to select multiple targets by "touching" each one of them (point-based selection), others allow the creation of a selection volume that collects all of the objects inside (volume-based selection). These different approaches indicate that there is still no consensus on which technique is the most efficient and effective, thus highlighting the importance of investigating the design space of multi-object acquisition in VR.

Fig. 1. Example scenarios of multi-object acquisition in VR. Left: a user is trying to select all the bottles in a virtual chemistry lab. Right: a user is attempting to acquire all the green data points in a data cloud for further investigation.

We highlight two design factors of VR multi-object acquisition that are still underexplored in the existing literature. First, while volume-based selection techniques have been proven to be useful in structured environments [16,28], further research is needed to verify their performance in more randomized scenarios. In these randomized scenarios, the position of targets does not follow a specific layout and is difficult to predict (e.g., Figure 1). Second, the effect of deselection, the mechanism that helps users remove unwanted selections, remains unclear in multi-object acquisition tasks. In such tasks where there are many interactable objects, no matter whether they are desired or undesired, it is likely that users will accidentally include unwanted objects in their initial selection. Thus, the deselection technique allows them to correct their errors. It may even enable new strategies for acquiring desired objects. For example, in a condition where the majority of the objects are targets, users may prefer selecting a whole volume of objects with a volume-based selection technique first and then remove unwanted objects with a deselection mechanism.

Our research, therefore, aims to explore how the choice of point-based and volume-based selection techniques, especially under randomized environments, and how the adoption of a deselection mechanism may affect user performance and experience. To achieve this, we first define a two-dimensional design space that considers point-/volume-based techniques and selection/deselection mechanisms. We implemented three techniques based on the design space and evaluated them in a user study that contains both randomized and structured scenarios. We found that a point-based technique was more robust in randomized scenarios, while a volume-based technique can perform particularly well in more structured, target-dense scenarios. We also found users rarely applied the deselection mechanism and mostly used them for error correction. Based on our findings, we provide implications that can inspire future designs of VR multi-object acquisition techniques.

2 Related Work

2.1 Object Acquisition in VR

Virtual pointing and virtual hand are the most common techniques for object acquisition in current VR systems [20,23,35,41]. Virtual pointing selects an object that intersects with a selection ray, while virtual hand selects an object that is "grabbed" by a user's hand. Many techniques have been proposed to overcome the limitation of virtual pointing and virtual hands (see reviews [1,14]). For example, because acquiring out-of-reach objects with a virtual hand is challenging, Poupyrev et al. [24] have proposed Go-Go which leverages a non-linear mapping function between real and virtual hands to extend the user's reach. More recent work proposed Ninja Hands which uses many hands to further extend the interactable area [26]. Baloup et al. [3] and Yu et al. [38] aimed to tackle the selection of small or even occluded targets, which are challenging scenarios for virtual pointing because of hand tremors and visual occlusion. Wagner et al. [32], on the other hand, tried to combine both virtual pointing and virtual hands to allow the technique to be adaptable to more complicated scenarios. However, most of these selection techniques were designed for selecting a single target, while being limited in selecting multiple targets as a user can only perform the selection serially (i.e., one target at a time).

2.2 Multi-Object Acquisition

Multi-object acquisition is a frequent task for both 2D interfaces (e.g., desktop, tablet) and 3D interfaces (e.g., VR). Dedicated techniques have been proposed in the literature for such a task, which can be roughly categorized into two main approaches: selecting targets point-by-point (serially, one-by-one) or group-by-group.

2D interfaces There are several common interaction scenarios that require tools for multi-object acquisition on 2D interfaces. A user may want to select a subset of items for coloring, transformation, removal, or other further operations. One option is to select the targets one by one [28] by holding an additional button (e.g., the Ctrl button). Another common option is to use a rectangle lasso (i.e., a selection volume) to select a group of targets. For example, users can select files by creating a rectangle area that encloses the desired files [34]. Mizobuchi and Yasumura's work evaluated the performance of lasso selection and found it was efficient in selecting targets that are aligned with high cohesiveness [21]. Another multi-object acquisition strategy is to automatically select similar objects altogether based on certain algorithms that calculate a similarity score (e.g., pixel-wise correlation) [8]. However, performance results from 2D multi-object acquisition techniques may not be easily transferred to 3D because of the additional depth dimension.

3D interfaces Lucas et al. were the first to present a taxonomy of 3D multi-object acquisition [16]. One design dimension was concurrency, which means that a technique can enable the selection of either one object per operation (serial selection) or a group of objects at once (parallel selection). The authors conducted a user study to evaluate serial and parallel selection techniques, and analyzed their selection performance within a structured environment where all objects were placed on a grid. The authors found that the parallel selection technique outperformed the serial technique when objects are placed on a highly cohesive layout and when the number of target objects is large.

Ulinski et al. [29,30] also investigated techniques that select a group of objects within a created box (cube). Specifically, three bimanual box creation methods were evaluated, which were Two Corners (TC), Hand-on-Corner (HOC), and Hand-in-Middle (HIM). The authors found that TC outperformed the two other methods regarding both completion time and accuracy but can introduce more arm strain. Other researchers have also investigated methods that create various shapes of selection volumes. For example, Kang et al.'s technique [13] allowed the creation of selection volumes with customizable shapes by performing different hand gestures. Although this seemed to provide a handy way of fitting the layout of the targets, it was shown to have lower efficiency when users wanted to adjust the volume to a desired shape, and it was difficult for users to remember the gesture set. Worm Selector [9] leveraged an easier method to create 3D volume by auto-connecting multiple 2D surfaces to form a 3D worm-like volume. Although this technique can be intuitive and efficient to use, it does not provide further support on editing the selected volumes.

Physically-Based Volumetric Selection [4] allowed users to manipulate the position of objects before making a selection so that they can cluster the desired objects first and then apply a selection volume. This mimics how people would manipulate multiple objects in the physical world. However, this technique lacks efficiency since it takes time to translate objects. Magic Wand [28] leveraged a grouping mechanism that automatically selects objects based on their local proximity to other objects. The technique was shown to be highly efficient and easy to use when the object layout followed certain structures. Slicing-Volume [22] used a tablet to map 3D data points onto 2D surfaces and enabled group-based selection by using a controller and a pen metaphor. Balloon [5] uses a 2D interface to control the selection in 3D environments through a balloon metaphor. The size of the position of the selection volume (i.e., the balloon) was controlled by a user's fingers.

Existing research has also investigated point-based selection techniques that can be applied for selecting multiple objects serially in 3D interfaces. Brushing mechanisms [12] select objects that intersect with the trajectory of the virtual hand or end-point of the ray. Sewing [15] selects an object by penetrating its surfaces which simulates the real-life action of sewing with a needle and fabric.

Summary We have identified two gaps in the literature. First, while most previous studies focus on applying volume-based selection techniques for more structured environments, it is unclear if they are still beneficial compared to point-based selection in randomized target layouts where the location of the targets can be difficult to predict [25,18]. Second, while previous research focuses on the selection phase of the acquisition process, the effect of object deselection (removing unwanted objects) has not been investigated. The ability of deselection can influence the overall selection strategies (e.g., selecting a large volume of objects that may include distractors and then trimming them off via deselection methods).

3 Technique Design

3.1 Design Space

We divided the design space into two dimensions. In the first design dimension, we separated a multi-object acquisition task into two action phases: selection and deselection. Selection represents the process of acquiring or identifying a particular subset of objects from the entire set of objects available [14]. Deselection represents the process of removing a subset of objects from the selected set of objects. In the second design dimension, we had point-based and volume-based techniques. A point-based method selects/deselects objects one by one while a volume-based method enables the selection/deselection of objects group by group. We used volume-based techniques because our literature review showed that it was the most prevalent for selecting multiple targets at once. Therefore, the design space leads to four combinations of techniques as shown in Figure 2. Among the four combinations, we deemed that point-based selection + volume-based deselection to be counter-intuitive because users rarely select a series of wrong objects individually and then deselect all of them. We thus removed it from further exploration.

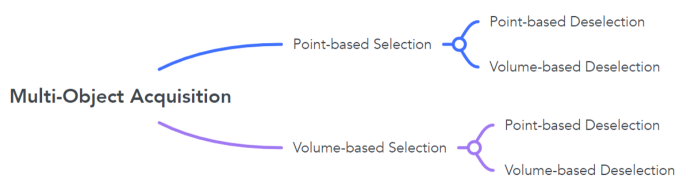

Fig. 2. Our design space resulted in four combinations of multi-object acquisition techniques.

3.2 Techniques

Based on the design space, we derived three techniques for multi-object acquisition. We made the following design factors consistent across all the techniques so that we could compare them through an experiment. All techniques leveraged game controllers rendered in the virtual environment as an indicator of the hand position. The techniques could be used with either one of the hands. There was consistent visual and audio feedback when object selection or deselection was triggered. Additionally, we rendered an outline to an object if it was selected and removed it once it was deselected.

PSPD (Point Selection + Point Deselection) With PSPD, a user can select an object by touching the object with the controller and pressing the trigger button (see Figure 3). A group of objects can be selected by triggering the selection of each object serially. One can deselect an object by touching the selected object and pressing the trigger. Deselection of a group of objects also happens serially. This technique is similar to virtual hands [23] but applied in the context of multi-object acquisition.

VSVD (Volume Selection + Volume Deselection) With VSVD, a user can select/deselect a group of objects with a volume-based selection mechanism. A user starts the process by creating a box-shape volume. Once the user presses the trigger of a controller, a box corner is initiated and fixed at the same position. The diagonal corner of the box updates as the position of the controller updates, so that the size

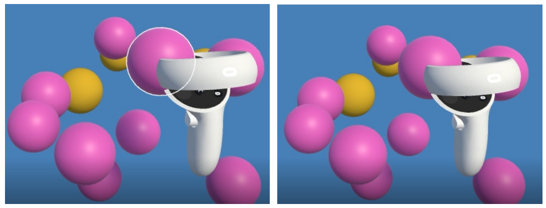

Fig. 3. With PSPD (Point Selection + Point Deselection), a user can select/deselect an object by touching the object with the controller and pressing the trigger button.

of the box increases or decreases based on a user's action (see Figure 4). Once the user releases the trigger, all objects within the box volume will be selected, and the selection box disappears. The same process is applied to the deselection phase.

A user can switch between selection and deselection mode by pressing the primary button of a controller. The mode was set to selection mode by default while entering a task. An indication of mode was attached to the center of the virtual controller, with a "+" sign representing the selection mode and the "-" representing the deselection mode. Unlike PSPD, we explicitly separated the two modes because we found it could cause user confusion if the two modes were combined during our pilot test.

We chose a box-shape selection volume to mimic the Two Corner technique proposed by Ulinski et al. [29,30], where the creation of a box was defined by the positions of two diagonal corners. However, unlike the Two Corner technique, which used both hands to manipulate the corners, our technique leveraged a single hand to create the cube to be consistent with the PSPD technique. The selection volume for both modes is semi-transparent, which was deemed as an effective representation of a fix-shaped selection box [37,40]. We distinguished the two modes further by using a green box to represent the selection box and a yellow box to represent the deselection box.

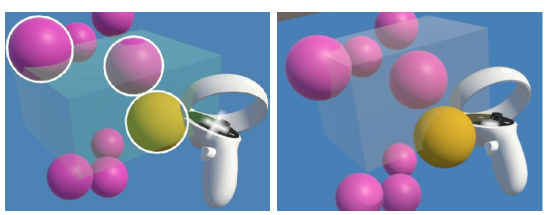

Fig. 4. With VSVD (Volume Selection + Volume Deselection), a user can select/deselect a group of object at once by creating a selection/deselection volume.

VSPD (Volume Selection + Point Deselection) With VSPD, a user can select a group of objects with a box-shape selection volume, like VSVD. The user can deselect one object at a time with the point-based mechanism, as in PSPD. The technique also has two explicit modes for selection and deselection, which requires a button click to switch between the different modes.

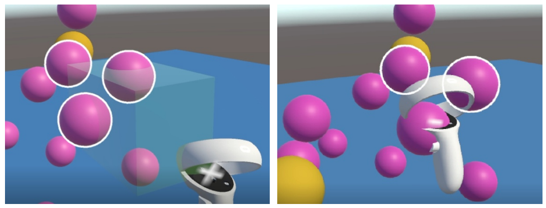

Fig. 5. With VSPD (Volume Selection + Point Deselection), a user can select a group of objects through a selection volume and deselect one object at a time.

4 User Study

4.1 Participants and Apparatus

We recruited 12 participants (5 women, 7 men) between the ages of 19 to 27 (mean = 22.83). Participants self-reported their familiarity with VR, and all of them had rather limited knowledge of VR (mean = 2.08, on a 5-point Likert scale). The system was implemented in Unity and the study was conducted on an Oculus Quest 2 headset.

4.2 Task Scenarios

We employed two types of task scenarios: randomized scenarios and structured scenarios. As we wanted to test whether volume-based selection would still be beneficial if the target layout was randomized and evaluate how selection and deselection mechanisms affect user performance, we leveraged a controlled scenario with randomly generated target layouts. Furthermore, as we also wanted to verify previous findings in more structured layouts, we also composed two structured scenarios.

For both task scenarios, there were targets and distractors, and all of them were selectable. While objects may be represented by different shapes in each task (e.g., a sphere, a book, molecules), the targets were always set in red to make them consistent across the tasks. All the tasks were in the primary working space [36] where objects were all in front of the user and were within arm-reach distance. We deemed that further manipulation of the objects normally happens within this space.

Randomized Scenarios In the randomized scenarios, all objects were generated in a random position within the region of 100cm×50cm×50cm. The objects were all spheres with a fixed diameter of 7.5cm, representing a relatively easier selection of each target as seen in previous work [2]. These scenarios were designed to assess the techniques' performance with more randomized target layouts, such as selecting some data points for further analysis in an immersive analytics scenario and acquiring certain items from a messy workspace, where the target positions may not follow a particular structure. Programmatically, we generated the objects with the built-in `Random.Range` function in Unity within a bounding box. The targets could overlap with each other. Since the overlapping could happen in all conditions, it is quite unlikely to be a confounding factor.

We varied two independent variables in the randomized scenarios: Task Difficulty and Target Majority. With Task Difficulty, we varied the total number of each scenario, with a total of 15 objects in the simple condition and a total of 40 in the difficult condition. The choice of these numbers was inspired by the result of a previous study [16], which showed that the volume-based selection techniques started to show a significant advantage over the point-based techniques when the target number was large (around 30). With Target Majority, we varied the proportion of the targets among all the objects. In the target minority condition, the total number of target objects was approximately 1/3 of the total number of objects, while in the target majority condition, the target objects were 2/3 of the total objects. With this variable, we aimed to investigate whether the proportion of the targets would affect a user's strategy of using selection and deselection methods.

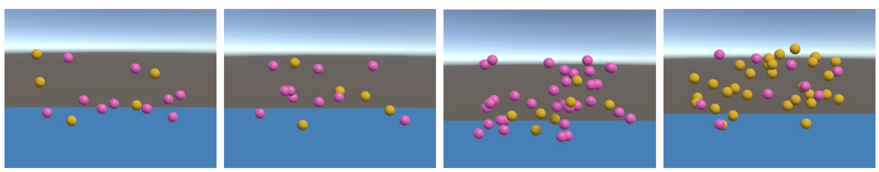

Fig. 6. Example conditions in the randomized scenarios.

Structured Scenarios We designed two application scenarios to verify findings from previous research that volume-based selection techniques may have a better performance when the layout of objects were more structured.

The first was a bookshelf scenario where participants were supposed to select books from a bookshelf. There were a total number of 30 books (24 targets), and all of them were in the size of 15cm×5cm×20cm. There were a total of 4 layers on the bookshelf, where each layer contained 7 or 8 books (see Figure 7, left). The targets were always next to each other and the selection of targets was predetermined. The second scenario was to select certain atoms from a molecule. There were 28 atoms in the molecule and all of them had a diameter of 7cm. There were 6 non-target atoms inside the molecule, which were placed together and were not in red. The two structured scenarios gave different layouts where all objects were placed at the same depth in the bookshelf scenario and the objects were placed at different depths in the molecule scenario. The parameters (i.e., target size, the number of targets) were fixed. The target number for both scenarios was larger than 20, with which the volume-based techniques were more likely to excel [16].

In sum, we studied 6 scenarios (= 4 randomized scenarios + 2 structured scenarios). The randomized scenarios were composed of 2 Task Difficulty × 2 Target Majority, whereas the structured scenario consisted of 1 bookshelf scenario (2D arrangement) and 1 molecule scenario (3D arrangement).

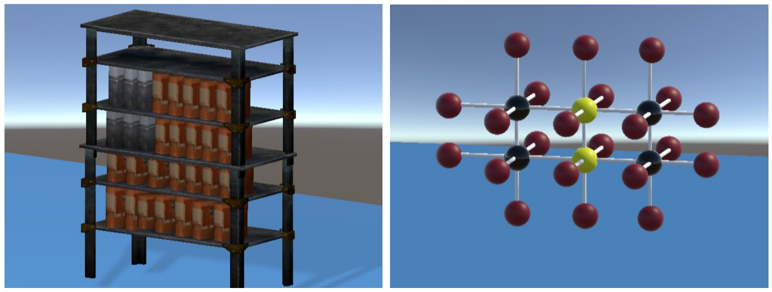

Fig. 7. Structured scenarios: the bookshelf scenario (left) and the molecule scenario (right).

4.3 Evaluation Metrics

Performance Measures To evaluate the techniques' performance, we employed the following two measures:

- *Selection time*: the time taken to complete the task for each trial.
- *Error rate*: the number of incorrect objects selected/not selected (i.e., false positives and false negatives) in relation to the total number of objects. The errors were recorded at the end of each trial, which was determined by participants pressing a finishing button on the controller.

Hand Movement Measures We were interested in investigating how the techniques influence hand movement distance, which can correlate with arm fatigue [36,39].

- *Hand movement distance*: the aggregation of the hand movement distance by accumulating the displacement of hand per frame.

Subjective Measures We further measured participants' subjective experience of using the techniques.

- *User Experience Questionnaires - short version (UEQ-S)* [27]: a questionnaire on 7-point scales regarding user experience.

Strategy Measures We were interested in participants' strategies to solve the multi-object acquisition task, particularly in volume-based selection methods (e.g., performing multiple separate selections or selecting a large volume of objects and then teasing out unwanted objects). The following two measures were used as indicators of users' strategies.

- *Number of operations*: the number of selections/deselections have been made for each task trial.
- *Objects per operation*: the averaged number of selected/deselected objects in each operation of selection or deselection. If there is no such operation in a trial, the value is set to zero.

4.4 Design and Procedure

The study employed a 3×6 (techniques × scenarios) within-subject design. The techniques and scenarios have been introduced in previous sections. For each condition, we provided 7 repetitive trials to balance the stability of the results and user engagement in the study. Different trials had different targets and positions (i.e., re-randomized). The 4 randomized scenarios and 3 techniques were counterbalanced for 12 participants, and the structured scenarios appeared after the randomized conditions.

Before the experiment, participants were briefly introduced to this study and were asked to fill out a pre-test demographic questionnaire. Participants were instructed on how to use the VR system and the techniques and had enough time to practice before the start of each session. We did not prime participants on different selection strategies to simulate how the techniques would be used in real-world scenarios. During the experiment, participants remained seated, with no physical obstacles within arm-reach distance. They were also asked to complete the task as quickly and accurately as possible. User feedback was collected with UEQ-S after participants completed a technique in each scenario. Participants were asked to take a break before starting a new technique. After they completed all the tasks, participants were asked to rank the techniques and provide oral feedback regarding the experience of using each technique. The experiment lasted around 45 minutes in total for each participant.

4.5 Hypotheses

We were specifically interested in verifying the following three hypotheses via our user study.

- *H1*. The point-based selection technique (PSPD) will outperform the volume-based selection techniques (VSVD and VSPD) in randomized scenarios where the majority of the objects are distractors. Because the number of targets is not high in those cases, selecting objects in groups may make less sense with a randomized layout.
- *H2*. The volume-based selection techniques (VSVD and VSPD) will outperform the point-based selection technique (PSPD) in randomized scenarios where the majority of the objects are target objects, as it might be easier for users to make a large group selection and remove unwanted objects rather than selecting the targets additively.
- *H3*. The volume-based selection techniques (VSVD and VSPD) will outperform the point-based selection technique (PSPD) in the structured scenarios because the two volume-based techniques can select multiple clustered objects at once.

5 Results

We collected 1512 trials of data (12 participants × 3 techniques × 6 scenarios × 7 repetition) from the within-subject experiment. The trials that had time and error rates greater than three deviations from the mean value had been discarded in each condition (30 trials, 3% in the randomized scenarios; 18 trials, 3.6% in the structured scenarios). After removing the outliers, we performed Shapiro-Wilk normality tests. The results indicated that the data were not normally distributed. Hence, we used

Aligned Rank Transform [33] to pre-process the data in both scenarios. We further conducted repeated-measures ANOVA (RM-ANOVA) and ART-based pairwise comparisons [10] to evaluate the performance and user experience of different techniques. In the following sections, we focus on main effects and two-way interaction effects that are related to the TECHNIQUE factor, because our main goal was to explore the difference in techniques under various conditions. We present the raw statistical results in the supplementary material.

5.1 Randomized Scenarios

Selection time RM-ANOVA showed that TECHNIQUE had a significant impact on task completion time ($F_{2,954}=103.50, p<.001, \eta_p^2=0.18$). On average, the point-based selection technique (PSPD) took the least amount of time while VSPD took the longest (see Figure 8 left). An interaction effect between TECHNIQUE and TARGET MAJORITY on complete time was identified ($F_{2,954}=18.98, p<.001, \eta_p^2=0.04$)—the influence of the majority of target-major or distractor-major on the two volume-based selection techniques (VSVD and VSPD) was significantly larger than on the point-based selection technique (PSPD), as shown in Figure 8 right. In the target minority scenario, the difference between VSVD and PSPD was not as large as in the target majority scenario.

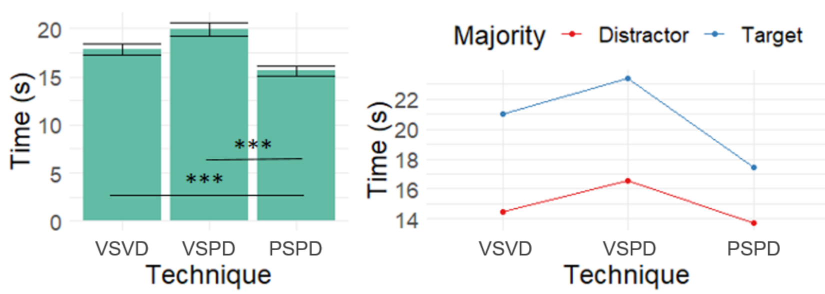

Fig. 8. Left: Average selection time ($\pm 1 S.E.$) of the techniques in the randomized scenarios. Right: interaction effect between Technique and Target Majority.

Error rate RM-ANOVA showed that TECHNIQUE had a significant main effect on error rate ($F_{2,954}=13.74, p<.001, \eta_p^2=0.03$). Interactions between TECHNIQUE × TARGET MAJORITY ($F_{2,954}=27.84, p<.001, \eta_p^2=0.06$) and TECHNIQUE × TASK DIFFICULTY ($F_{2,954}=18.06, p<.001, \eta_p^2=0.04$) were also identified. Pairwise comparisons showed that PSPD had a lower error rate than VSVD ($p=.024$), while VSVD led to lower error rate than VSPD ($p=.031$). However, the overall error rates of the techniques were very low (PSPD: 0.97% ; VSVD: 1.15%; VSPD: 1.19%;).

Hand movement distance RM-ANOVA indicated that TECHNIQUE had a significant main effect on hand movement distance ($F_{2,954}=56.84, p<.001, \eta_p^2=0.11$). As in Figure 9 left, the two volume-based techniques led to longer hand movement distance than the point-based technique (both $p < .001$). An interaction between TECHNIQUE and TARGET MAJORITY was also identified ($F_{2,954} = 8.79, p < .001, \eta_p^2 = 0.02$). While the hand movement distance increased when the majority of the objects became targets, as indicated in Figure 9 right, the influence of TARGET MAJORITY was much larger on the volume-based techniques than on the point-based technique. We also observed that the hand movement distance of the three techniques did not differ significantly in the target minority condition.

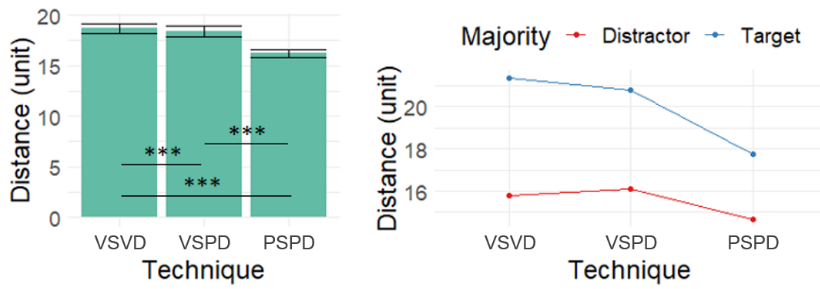

Fig. 9. Left: Average hand movement distance ($\pm 1 S.E.$) of the techniques in the randomized scenarios. Right: an interaction effect between Technique and Target Majority.

Number of operations For the number of selection operations, RM-ANOVA indicated that TECHNIQUE had a significant main effect ($F_{2,954} = 797.99, p < .001, \eta_p^2 = 0.63$). Interaction effects of TECHNIQUE × TARGET MAJORITY ($F_{2,954} = 268.29, p < .001, \eta_p^2 = 0.36$) and TECHNIQUE × TASK DIFFICULTY ($F_{2,954} = 518.24, p < .001, \eta_p^2 = 0.52$) were also identified. As shown in Figure 10 left, the point-based technique (PSPD) required significantly more operations than the other two volume-based techniques (VSVD and VSPD) when the target number was high in the difficult condition. Also, the number of selection operations of the three techniques was quite similar in the other three conditions.

For the number of deselection operations, RM-ANOVA suggested that TECHNIQUE had a significant main effect ($F_{2,954} = 98.30, p < .001, \eta_p^2 = 0.17$). Interaction effects of TECHNIQUE × TARGET MAJORITY ($F_{2,954} = 4.10, p = .017, \eta_p^2 = 0.01$) and TECHNIQUE × TASK DIFFICULTY ($F_{2,954} = 42.97, p < .001, \eta_p^2 = 0.08$) were also identified. Overall, as shown in Figure 10 right, the number of deselection operations was quite low (i.e., less than 1.15 on average). The number for the difficulty-target major condition was slightly higher for all techniques than for the other conditions.

Objects per operation For the number of objects per selection operation, RM-ANOVA indicated that TECHNIQUE had a significant main effect ($F_{2,954} = 438.78, p < .001, \eta_p^2 = 0.48$). Interaction effects of TECHNIQUE × TARGET MAJORITY ($F_{2,954} = 154.05, p < .001, \eta_p^2 = 0.24$) and TECHNIQUE × TASK DIFFICULTY ($F_{2,954} = 130.46, \eta_p^2 = 0.21$) were also identified. As shown in Figure 11, the number of selected object per

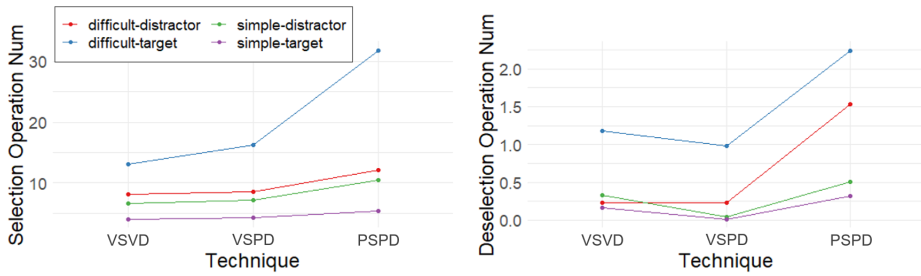

Fig. 10. Average number of (de-)selection operations in the randomized conditions.

operation was higher in the difficult-target majority condition for both volume-based selection techniques (VSVD and VSPD).

For the number of objects per deselection operation, RM-ANOVA indicated that Technique had a significant main effect ($F_{2,954}=77.84, p<.001, \eta_p^2=0.14$). An interaction effect of Technique × Task Difficulty ($F_{2,954}=25.12, p<.001, \eta_p^2=0.05$) was also identified. Very limited objects were getting deselected per operation (i.e., lower than 1 on average) for all the techniques.

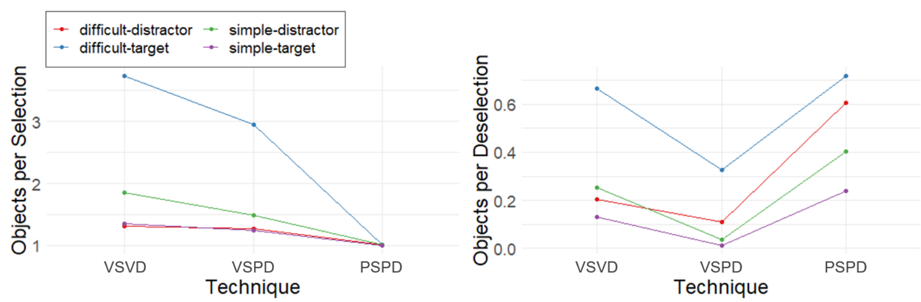

Fig. 11. Average number of (de-)selected objects per operation in the randomized conditions.

Table 1. Summary of the statistical main effects of Technique on the evaluation metrics in the structured scenarios.

Dependent Variable	**Bookshelf** *F-value*	*p-value*	η_p^2	**Molecule** *F-value*	*p-value*	η_p^2
Selection time	38.06	<.001	0.25	1.50	0.224	0.01
Error rate	3.35	0.037	0.03	1.74	0.176	0.02
Distance	14.92	<.001	0.11	3.94	0.021	0.03
Selection operation num	378.28	<.001	0.77	325.02	<.001	0.74
Deselection operation num	66.52	<.001	0.37	5.54	.004	0.05
Objects per selection	436.07	<.001	0.79	376.65	<.001	0.77
Objects per deselection	36.91	<.001	0.24	4.09	.018	0.03

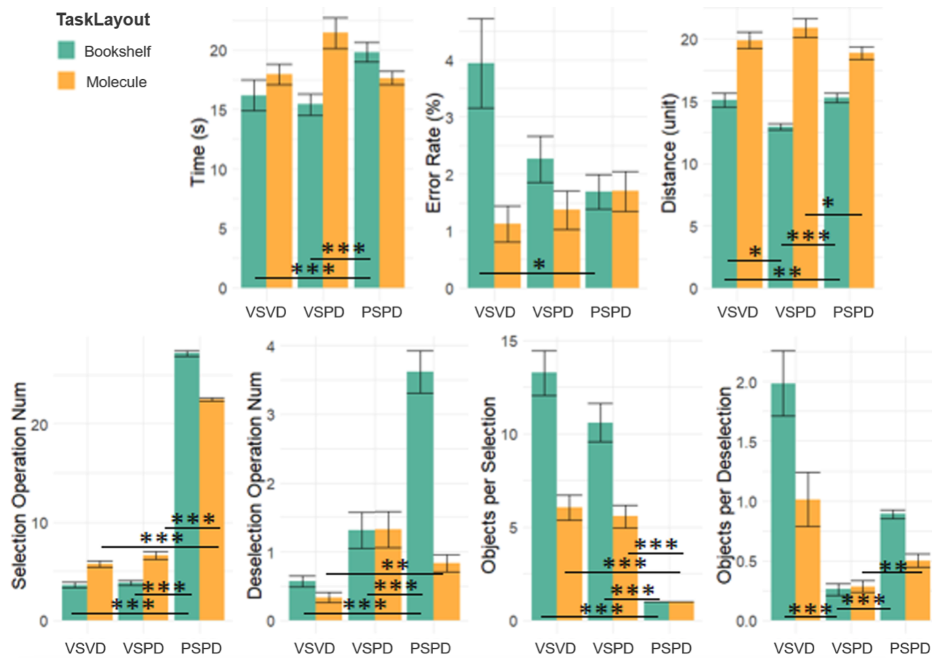

Fig. 12. Average selection time, error rate, hand movement distance, number of selection/deselection operations, and number of selected/deselected object per operation for both of the structured scenarios. Error bar indicates one standard error.

5.2 Structured Scenarios

We summarize the statistical main effects of TECHNIQUE on different evaluation metrics in Table 1 and the average results of the techniques in Figure 12. More details on the ART-based RM-ANOVAs and pairwise comparisons can be found in the provided supplementary material.

In the bookshelf scenario, overall, the volume-based techniques (VSVD and VSPD) were significantly faster than the point-based technique (PSPD), while all of them maintained a relatively low error rate ($<4\%$) throughout. VSPD led to the lowest movement distance, while PSPD was the highest. The selection operations were significantly lower for the volume-based techniques (VSVD and VSPD) than PSPD, and correspondingly, the objects per selection of these techniques were significantly higher. The number of deselection operations was significantly lower for VSVD and VSPD. The objects per deselection of VSPD was the lowest while VSVD was the highest.

In the molecule scenario, overall, the performance difference in terms of selection time and error rate among the three techniques was not identified to be significant. However, the hand movement distance while using VSPD was found to be significantly higher than PSPD. Similar to the bookshelf scenario, we also found that the volume-based techniques led to a much lower number of selection operations and a higher object number per selection. VSVD led to significantly lower deselection operations than PSPD, and VSPD led to a significantly lower number of objects per deselection as compared to PSPD. We infer group selection led to higher accuracy during the selection phase in structured scenarios, so there was less need to deselect objects.

5.3 User Experience

The UEQ-S results are summarized in Table 2. Overall, the PSPD led to the highest pragmatic value, while the VSVD and VSPD were more interesting to use. VSVD led to slightly higher overall quality than PSPD.

Table 2. UEQ-S results of each technique (the higher the better).

Technique	Pragmatic	Hedonic	Overall
PSPD	2.31 (Excellent)	-0.04 (Bad)	1.14 (> Average)
VSVD	1.35 (> Average)	1.13 (> Average)	1.24 (> Average)
VSPD	1.18 (> Average)	0.5 (< Average)	0.84 (< Average)

5.4 Observations and Subjective Feedback

We observed that most participants tended to select the objects one by one rather than creating selection volumes in the randomized environment even with the volume-based selection techniques. One participant mentioned the reason behind this behavior: the volume-based selection techniques were used as the point-based selection technique by instantiating very small cubes, and that was an easier way for this participant to complete the task in the randomized scenarios. In fact, seven of the participants self-reported that they preferred to make multiple separated selection operations rather than creating a large selection volume and then make deselections to remove unwanted objects. Half of the participants did not realize the later strategy, and the others felt the interaction of deselection to be overly complicated in randomized scenarios.

We also observed that participants moved the controller forward and backward repetitively to try to include or remove objects when creating a selection volume. Two of the participants commented that they had to make a few attempts to resize the selection volume to know how far they have reached.

When asking participants to rank the techniques, PSPD was the most favorite in general (10/12). Most preferred to use the volume-based selection techniques in the structured scenarios and the point-based selection technique in the randomized scenarios. Two of them further commented that VSVD was the most useful in the structured scenarios since it could cancel the selection of more than one object per operation.

6 Discussion

In the following, we compare and discuss the techniques based on our design dimensions: point- vs. volume-based technique and selection vs. deselection. We also examine the effect of different environmental factors (i.e., randomized vs. structured scenarios, task difficulties, and target/distractor majority) on technique performance.

6.1 Point- vs. Volume-based Selection

Overall, when the target locations did not follow a certain layout (i.e., in the randomized scenarios), we found that the point-based technique (PSPD) led to significantly lower selection time and shorter movement distance as compared to the volume-based techniques (VSVD and VSPD). The difference between the techniques was larger in scenarios where the majority of the objects were targets. This contradicted our initial hypothesis (*H2*) that the volume-based techniques would outperform the point-based technique because of their convenience in selecting a large group of objects. *H1* was partially supported because the point-based technique led to, on average, lower selection time when the majority of the objects were distractors.

We identified two potential reasons why the volume-based techniques led to worse performance than the point-based technique in randomized scenarios. First, volume-based selection techniques might not be suitable for randomized scenarios. Indeed, both participants' comments and selected objects per operation indicated that they preferred to use the volume-based techniques similar to how they use the point-based technique by making small, additive selections (i.e., selecting a small number of objects per operation). It might also be the chaotic nature of the randomized environment that made the participants believe that it might not be beneficial to select a large group of objects altogether.

Second, based on our observation and participants' feedback, the difficulty of depth perception might also be one reason for delaying the task completion of the volume-based techniques. For example, participants were found to move the controller forward and backward repetitively to try to gauge how far they have reached and whether they would be able to include or remove particular objects. These findings are in line with Ulinski et al.'s work [29,30], where participants also had difficulty in creating volumes since the depth information was not clear enough in a monitor. Our immersive 3D VR setting, which allowed participants to observe the selection volumes from different perspectives, did not seem to alleviate this issue of depth perception. Notably, we have provided proper depth cues for objects in the 3D environment. For example, strong depth cues like shading, occlusion, and relative object sizes were given. One issue might be because the selection volume was rendered in a monotonous color [19,31].

In the (structured) bookshelf scenario, both volume-based techniques outperformed the point-based technique in terms of selection time while maintaining a relatively low error rate ($<4\%$). However, in the (structured) molecule scenario, the performance difference among the three techniques was not significant. In both cases, the volume-based techniques led to a significantly lower number of selection operations and thus higher object number per selection. These findings partially confirmed *H3* and indicated that participants were more willing to apply the group-based selection mechanism when objects were structured (e.g., on average $>$10 objects per selection in the bookshelf scenario and $>$5 objects per selection in the molecule scenario).

In both scenarios, the volume-based selection techniques did result in less number of selection operations (thus higher object number per selection) in a more complex condition (i.e., difficult-target majority in the randomized scenarios). This indicates that volume-based selection activities did happen when the targets were clustered together, but it was not more efficient to do so.

6.2 Selection vs. Deselection

According to the selection and deselection operation numbers, we found that, overall, participants applied selection more frequently while rarely using the deselection mechanism. We also found that few participants seek to use a new strategy multi-object acquisition as proposed in *H2*, where we assumed that users would select a large group of objects and then remove unwanted objects. They preferred selecting objects additively by making several separate selections, even with the volume-based selection techniques.

6.3 The Effect of Environmental Factors

We found changing Task Difficulty in randomized scenarios led to a somewhat more consistent impact on the performance of different techniques (all of them suffered similarly in more challenging scenarios [1,14]). In contrast, when switching from distractor majority to target majority conditions, both volume-based techniques were more severely impacted as compared to the point-based technique—there was a more significant increase in selection time and hand movement distance for the volume-based techniques.

Overall, our results suggest that the point-based technique (i.e., PSPD) was more robust under different environmental factors which resulted in its highest pragmatic value in UEQ-S [27]. The volume-based technique (i.e., VSVD and VSPD) can reduce the number of operations across different environments, and may perform especially well in more structured, target-dense scenarios.

6.4 Design Implications

Based on the study results and our discussions, we distilled the following design implications.

- When the target locations do not follow a specific layout (i.e., with more randomized, hard-to-predict object locations), we recommend choosing the point-based selection technique (PSPD) among the three techniques (PSPD, VSVD, and VSPD). Our results suggest that the point-based technique (PSPD) is not only more efficient than the volume-based techniques (VSVD and VSPD) but is also more robust across different environmental conditions (e.g., the majority of objects are targets or distractors).
- When the targets are structured or clustered, we recommend considering applying a volume-based technique (either VSVD or VSPD). Our results suggest that VSVD and VSPD tend to perform well on structured layouts. Furthermore, VSVD and VSPD led to a significantly lower number of selection operations in those scenarios.
- We recommend enabling the deselection mechanism (i.e., allowing users to remove unwanted selections). However, be mindful that users may only use this function occasionally to correct their errors and may not switch their acquisition strategies depending on enabling/disabling this function.

– To further improve our volume-based selection techniques (VSVD and VSPD), we recommend focusing on providing a better depth perception of the selection volume. Unlike 2D group-based selection techniques used in computers (i.e., a 2D rectangle selection volume), in a 3D VR environment, we might want to apply extensive depth cues for the selection volume itself (not a transparent box with a monotonous color). Providing additional depth cues like color/texture gradients is likely to be beneficial.

6.5 Limitations and Future Work

We have identified several limitations of this research. First, we only evaluated one type of volume-based selection mechanism (i.e., a cube selection volume [29,30]), which seemed to be a good approximation of the rectangle lasso selection in PCs. We acknowledge that there are other types of strategies in VR, like using spherical selection volume or applying bimanual selection volume creation methods [25]. The choice of the volume-based selection mechanism can influence the technique's performance.

Second, we only evaluated one parameter level of the structured scenarios (i.e., the bookshelf and molecule scenarios). We did this because we wanted to focus on the randomized scenarios as they were less explored in previous research. While we were interested in verifying previous findings on structured scenarios, we did not want to put too much emphasis on these conditions considering the workload of the participants according to our pilot study (the experiment already lasted about 45 minutes). Also concerning the participant workload, we did not examine the effect of other factors such as target disparity, spread of targets, and target intersection and occlusion [1,6,11,38] while kept them the same across conditions. Future research may scrutinize their effects and potential interactions in different application scenarios.

Third, all the participants in this study were novice users that only had limited experience in using VR. Future studies could include expert users to reduce the potential novelty effect [6]. Also, expert users who have experienced other off-the-shelf multiple interaction techniques in VR might provide interesting insights.

7 Conclusion

In this research, we investigated point-based and volume-based multi-object acquisition in VR. Specifically, we leveraged a two-dimensional design space that considered (1) point- or volume-based operation and (2) selection and deselection mechanisms. We derived three techniques based on the design space, and explored the performance and usability of these techniques under both randomized and structured scenarios. We found that point-based techniques are more efficient in randomized scenarios, while volume-based techniques are better suited for structured scenarios. We also found that users rarely applied the deselection mechanism and mostly used them for error correction. Based on our results, we discuss potential reasons behind technique performance under different conditions, and provided design implications that can aid the future design of similar techniques. We believe this research provides valuable lessons for designers to both optimize their interaction techniques and adopt deselection mechanisms in VR multi-object acquisition tasks.

References

1. Argelaguet, F., Andujar, C.: A survey of 3d object selection techniques for virtual environments. Computers & Graphics **37**(3), 121–136 (2013). https://doi.org/10.1016/j.cag.2012.12.003
2. Ariza, O., Bruder, G., Katzakis, N., Steinicke, F.: Analysis of proximity-based multimodal feedback for 3d selection in immersive virtual environments. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). pp. 327–334. IEEE (2018). https://doi.org/10.1109/VR.2018.8446317
3. Baloup, M., Pietrzak, T., Casiez, G.: Raycursor: A 3d pointing facilitation technique based on raycasting. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. p. 1–12. CHI '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3290605.3300331
4. Benavides, A., Khadka, R., Banic, A.: Physically-Based Bimanual Volumetric Selection for Immersive Visualizations. In: Chen, J.Y., Fragomeni, G. (eds.) Virtual, Augmented and Mixed Reality. Multimodal Interaction, vol. 11574, pp. 183–195. Springer International Publishing, Cham (2019), series Title: Lecture Notes in Computer Science
5. Benko, H., Feiner, S.: Balloon Selection: A Multi-Finger Technique for Accurate Low-Fatigue 3D Selection. In: 2007 IEEE Symposium on 3D User Interfaces. p. 4142849. IEEE, Charlotte, NC, USA (2007). https://doi.org/10.1109/3DUI.2007.340778
6. Bergström, J., Dalsgaard, T.S., Alexander, J., Hornbæk, K.: 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. ACM, Yokohama Japan (May 2021). https://doi.org/10.1145/3411764.3445193
7. Bowman, D.A. (ed.): 3D user interfaces: theory and practice. Addison-Wesley, Boston (2005)
8. Dehmeshki, H., Stuerzlinger, W.: Design and evaluation of a perceptual-based object group selection technique. In: Proceedings of the 24th BCS Interaction Specialist Group Conference. p. 365–373. BCS '10, BCS Learning amp; Development Ltd., Swindon, GBR (2010)
9. Dubois, E., Hamelin, A.: Worm Selector: Volume Selection in a 3D Point Cloud Through Adaptive Modelling. International Journal of Virtual Reality **17**(1), 1–20 (Jan 2017). https://doi.org/10.20870/IJVR.2017.17.1.2884
10. Elkin, L.A., Kay, M., Higgins, J.J., Wobbrock, J.O.: An aligned rank transform procedure for multifactor contrast tests. In: The 34th Annual ACM Symposium on User Interface Software and Technology. pp. 754–768 (2021). https://doi.org/10.1145/3472749.3474784
11. Elmqvist, N., Tsigas, P.: A taxonomy of 3d occlusion management techniques. In: 2007 IEEE Virtual Reality Conference. pp. 51–58. IEEE (2007). https://doi.org/10.1109/VR.2007.352463
12. Hurter, C., Riche, N.H., Drucker, S.M., Cordeil, M., Alligier, R., Vuillemot, R.: FiberClay: Sculpting Three Dimensional Trajectories to Reveal Structural Insights. IEEE Transactions on Visualization and Computer Graphics **25**(1), 704–714 (Jan 2019). https://doi.org/10.1109/TVCG.2018.2865191
13. Kang, J., Zhong, K., Qin, S., Wang, H., Wright, D.: Instant 3D design concept generation and visualization by real-time hand gesture recognition. Computers in Industry **64**(7), 785–797 (Sep 2013). https://doi.org/10.1016/j.compind.2013.04.012
14. LaViola Jr, J.J., Kruijff, E., McMahan, R.P., Bowman, D., Poupyrev, I.P.: 3D user interfaces: theory and practice. Addison-Wesley Professional (2017)
15. Li, Y., Sarcar, S., Kim, K., Tu, H., Ren, X.: Designing successive target selection in virtual reality via penetrating the intangible interface with handheld controllers.

International Journal of Human-Computer Studies **165**, 102835 (Sep 2022). https://doi.org/10.1016/j.ijhcs.2022.102835
16. Lucas, J., Bowman, D., Chen, J., Wingrave, C.: Design and Evaluation of 3D Multiple Object Selection Techniques (Jan 2005)
17. M. Göbel: Industrial Applications of VEs. IEEE Computer Graphics and Applications **16**(01), 10–13 (Jan 1996). https://doi.org/10.1109/MCG.1996.10005
18. Marriott, K., Schreiber, F., Dwyer, T., Klein, K., Riche, N.H., Itoh, T., Stuerzlinger, W., Thomas, B.H.: Immersive analytics, vol. 11190. Springer (2018)
19. Matatko, A., Bollmann, J., Müller, A.: Depth perception in virtual reality. Advances in 3D Geo-Information Sciences pp. 115–129 (2011). https://doi.org/https://doi.org/10.1007/978-3-642-12670-3_7
20. Mine, M.R.: Virtual Environment Interaction Techniques. Tech. rep., University of North Carolina at Chapel Hill, USA (1995)
21. Mizobuchi, S., Yasumura, M.: Tapping vs. circling selections on pen-based devices: Evidence for different performance-shaping factors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. p. 607–614. CHI '04, Association for Computing Machinery, New York, NY, USA (2004). https://doi.org/10.1145/985692.985769
22. Montano-Murillo, R.A., Nguyen, C., Kazi, R.H., Subramanian, S., DiVerdi, S., Martinez-Plasencia, D.: 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. IEEE, Atlanta, GA, USA (Mar 2020). https://doi.org/10.1109/VR46266.2020.00023
23. Poupyrev, I., Ichikawa, T., Weghorst, S., Billinghurst, M.: Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. Computer Graphics Forum **17**(3), 41–52 (Aug 1998). https://doi.org/10.1111/1467-8659.00252
24. Poupyrev, I., Billinghurst, M., Weghorst, S., Ichikawa, T.: 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 - UIST '96. pp. 79–80. ACM Press, Seattle, Washington, United States (1996). https://doi.org/10.1145/237091.237102
25. Romagnoli, G.: Best multi-object selection methods for virtual reality (2021). https://doi.org/https://bootcamp.uxdesign.cc/best-multi-object-selection-methods-for-virtual-reality-97027160f584
26. Schjerlund, J., Hornbæk, K., Bergström, J.: Ninja hands: Using many hands to improve target selection in vr. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3411764.3445759
27. Schrepp, M., Hinderks, A., Thomaschewski, J.: Design and evaluation of a short version of the user experience questionnaire (ueq-s). International Journal of Interactive Multimedia and Artificial Intelligence, 4 (6), 103-108. (2017). https://doi.org/10.9781/ijimai.2017.09.001
28. Stenholt, R.: Efficient selection of multiple objects on a large scale. In: Proceedings of the 18th ACM symposium on Virtual reality software and technology - VRST '12. p. 105. ACM Press, Toronto, Ontario, Canada (2012). https://doi.org/10.1145/2407336.2407357
29. Ulinski, A., Wartell, Z., Hodges, L.F.: Bimanual task division preferences for volume selection. In: Proceedings of the 2007 ACM symposium on Virtual reality software and technology - VRST '07. p. 217. ACM Press, Newport Beach, California (2007). https://doi.org/10.1145/1315184.1315228
30. Ulinski, A., Zanbaka, C., Wartell, Z., Goolkasian, P., Hodges, L.F.: Two Handed Selection Techniques for Volumetric Data. In: 2007 IEEE Symposium on 3D User Interfaces. p. 4142853. IEEE, Charlotte, NC, USA (Mar 2007). https://doi.org/10.1109/3DUI.2007.340782

31. Vienne, C., Masfrand, S., Bourdin, C., Vercher, J.L.: Depth perception in virtual reality systems: effect of screen distance, environment richness and display factors. IEEE Access **8**, 29099–29110 (2020). https://doi.org/10.1109/ACCESS.2020.2972122
32. Wagner, J., Stuerzlinger, W., Nedel, L.: Comparing and Combining Virtual Hand and Virtual Ray Pointer Interactions for Data Manipulation in Immersive Analytics. IEEE Transactions on Visualization and Computer Graphics **27**(5), 2513–2523 (May 2021). https://doi.org/10.1109/TVCG.2021.3067759
33. Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 143–146. ACM, Vancouver BC Canada (May 2011). https://doi.org/10.1145/1978942.1978963
34. Yamanaka, S., Stuerzlinger, W.: Modeling fully and partially constrained lasso movements in a grid of icons. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. p. 1–12. CHI '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3290605.3300350
35. Yu, D., Liang, H.N., Lu, F., Nanjappan, V., Papangelis, K., Wang, W., et al.: Target selection in head-mounted display virtual reality environments. J. Univers. Comput. Sci. **24**(9), 1217–1243 (2018). https://doi.org/https://doi.org/10.3217/jucs-024-09-1217
36. Yu, D., Lu, X., Shi, R., Liang, H.N., Dingler, T., Velloso, E., Goncalves, J.: Gaze-supported 3d object manipulation in virtual reality. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. pp. 1–13 (2021). https://doi.org/10.1145/3411764.3445343
37. Yu, D., Zhou, Q., Dingler, T., Velloso, E., Goncalves, J.: Blending on-body and mid-air interaction in virtual reality. In: 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). pp. 637–646. IEEE (2022). https://doi.org/10.1109/ISMAR55827.2022.00081
38. Yu, D., Zhou, Q., Newn, J., Dingler, T., Velloso, E., Goncalves, J.: Fully-occluded target selection in virtual reality. IEEE Transactions on Visualization and Computer Graphics **26**(12), 3402–3413 (2020). https://doi.org/10.1109/TVCG.2020.3023606
39. Yu, D., Zhou, Q., Tag, B., Dingler, T., Velloso, E., Goncalves, J.: Engaging participants during selection studies in virtual reality. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). pp. 500–509 (2020). https://doi.org/10.1109/VR46266.2020.00071
40. Zhai, S.: Human performance in six degree of freedom input control [microform]. (Jan 1995)
41. Zhou, Q., Yu, D., Reinoso, M.N., Newn, J., Goncalves, J., Velloso, E.: Eyes-free target acquisition during walking in immersive mixed reality. IEEE Transactions on Visualization and Computer Graphics **26**(12), 3423–3433 (2020). https://doi.org/10.1109/TVCG.2020.3023570