MOA: Efficient Scene-aware Multi-object Arrangement in VR Supplementary Material

Xuehuai Shi, Yuhan Duan, Ziteng Wang, Jian Wu, Zhiwen Shao, Jieming Yin, and Lili Wang

In this document, we provide the details of the pilot user study, user studies 1 and 2 in support of the main text.

1 Pilot User Study: Coefficients Optimization

In this section, we conduct a pilot user study to optimize the MOA coefficients, thereby improving MOA's task performance in general multi-object arrangement scenes. MOA consists of two steps: MOA_s and MOA_m. There are three coefficients in MOA: the importance coefficient α of MOA_s, the structural quantity coefficient β, and the element thickness coefficient γ of MOA_m. In the multi-object arrangement task, MOA first executes MOA_s for initial selection, then proceeds to later manipulation using MOA_m, with the efficiency of these two steps being independent of each other. Therefore, we first assess the impact of α in MOA_s to find the optimal value; subsequently, we evaluate the effects of β and γ in MOA_m to determine their optimal values.

We formulate two hypotheses for the pilot user study:

H1. Different α values used in MOA significantly affect the task performance of the initial selection step in the general multi-object arrangement scene.

H2. Different β and γ values used in MOA significantly affect the task performance of the later manipulation step in the general multi-object arrangement scene.

1.1 User Study Design

Apparatus. Our system uses a PICO 4 Pro HMD powered by a workstation with a 3.8GHz Intel(R) Core(TM) i7-10700KF CPU, 32GB of RAM, an NVIDIA GeForce GTX 3080Ti graphics card, and an HTC Vive tracker. The resolution of the HMD is 2160×2160 pixels for each eye, and the field-of-view is 105°. The whole system was running at 90 fps for each eye. We use the built-in programs of PICO 4 Pro to implement gaze and hand gesture tracking. Our program is developed with C# and HLSL, and is run in Unity 2021.3.8f1.

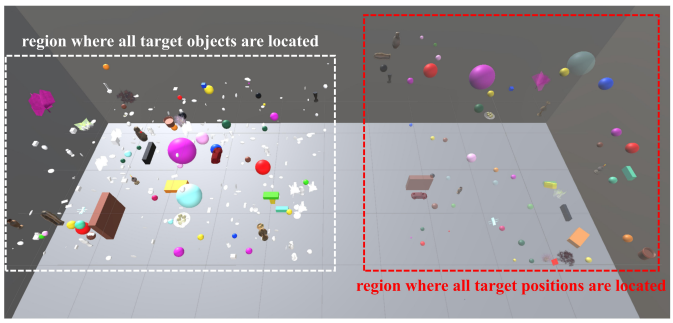


Fig. 1: Visualization of *general scene*.

Test Scene Construction. To fully simulate the complex interference and occlusion scenarios encountered during general multi-object arrangement tasks in VR, we carefully design a highly occluded general testing scene, *general scene*, based on previous research, as shown in Fig. 1. We determine the number of objects according to the multi-object manipulation scene arrangement scheme in Maslych et al. [1], while the object shapes and target position layouts are based on findings from the multi-object arrangement review by Bergstrom et al. [2]. Following Maslych et al. [1], we set the number of candidate and target objects to the median values across all scenes, 256 and 60 respectively, and we place them randomly within the current field of view. According to Bergstrom et al. [2], 55.6% of candidate objects are spherical, 22.2% cubic, and 22.2% real-world objects, which have complex shapes such as toys and sculptures. Regarding target position layout, 40.0% are randomly distributed, 36.0% arranged in a circular layout, and 24.0% organized in a grid layout [2].

All objects and target positions in the *general scene* are fixed and static. Since the experiments in this paper aim to compare the performance and user experience across different method conditions in multi-object arrangement tasks, fixing object locations and target positions effectively eliminates distractions caused by dynamic factors and prevents non-reproducibility in experimental results. This ensures the reliability of the results when evaluating condition effects in multi-object manipulation tasks, thereby enhancing the comparability and interpretability of the data. Therefore, all test scenes in this paper have fixed object locations and target positions by design.

Participants. We recruit 16 participants, including nine males and seven females, aged from 18 to 50, with an average age of 27. All participants have normal vision or corrected-to-normal vision. Ten of them have experience using HMD VR applications before the study.

Conditions. The proposed method MOA is defined as $MOA_s(\alpha) + MOA_m(\beta,\gamma)$, where different coefficient values affect the multi-object arrangement task performance. Below, we give the coefficient levels.

- In initial selection, $\alpha \in \{1/2, 2/3, 3/4, 1\}$ in MOA_s has four levels. Values of $\alpha < 1$ mix historical and current-round importance. The value of α ranges from 0 to 1. A lower α emphasizes historical importance, while a higher α emphasizes current-round importance. $\alpha = 1$ condition represents using only the current round's importance, serving as a baseline where historical importance is entirely discarded.
- In later manipulation, $\beta \in \{0, 6, 8, 10, +\infty\}$ in MOA_m has five levels, while $\gamma \in \{3, 6, 9\}$ has three levels. The coefficient β defines the maximum number of guiding elements in Ω. When $\beta = 0$, MOA_m does not use Ω to guide the later manipulation; when $\beta = +\infty$, MOA_m imposes no limit on the number of elements in Ω. The coefficient γ controls the thickness of all guiding elements in Ω. The greater γ, the thinner the guiding elements become.

There are a total of 19 conditions in this user study, i.e., $MOA_s(\alpha \in \{1/2, 2/3, 3/4, 1\}) + MOA_m(\beta \in \{0, 6, 8, 10, +\infty\}, \gamma \in \{3, 6, 9\})$.

Task and Procedure. This study involves two tasks: task 1 requires selecting all target objects from candidate objects using MOA_s, and task 2 involves manipulating all target objects to designated target positions using MOA_m. To maintain high engagement and retention rates for participants throughout the experiment, this study requires participants to complete the experiment over five consecutive days, ensuring that each experimental session lasts no more than 30 minutes [3]. Task 1 is completed on the first day. Before the formal experiment begins, participants spend 30 seconds using $MOA_s(\alpha = 3/4)$ for a brief practice session. In the subsequent formal experiment, each participant needs to complete 4 trials, with each trial corresponding to one α level in

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

$MOA_s(\alpha)$. To balance potential learning effects, the presentation order of these four conditions is fully balanced across all 16 participants using a 4×4 Latin Square design.

Task 2 is distributed over the second to fifth days. Before starting this task on the second day, participants spend another 30 seconds practicing with $MOA_m(\beta=8,\gamma=6)$. This task requires participants to complete all 15 trials, which correspond to all level combinations of β and γ in $MOA_m(\beta,\gamma)$. Considering that the number of conditions (15) makes a fully balanced design infeasible, we adopt a randomization approach: a unique trial order is generated for each participant to mitigate sequence effects and fatigue effects. Participants complete 4, 4, 4, and 3 trials respectively on the following four days, according to their individual randomized order.

For both tasks, to ensure fair comparisons, the x,y coordinates of initial positions for all conditions are fixed, the HMD is placed on the ground to ensure the participant's viewpoint height when wearing the HMD matches their actual height, and all candidate objects and target locations remain within the participants' initial field of view [2]. Each participant completes 19 trials in total. A total of 16 (participants) × 19 (conditions) = 304 trials are collected.

Metrics. Since the interaction mode of $MOA_s(\alpha)+MOA_m(\beta,\gamma)$ remains consistent across different coefficient levels, the effects on task load and convenience are uniform within the same unit of interaction time. Therefore, this study evaluates only the task performance of tasks 1 and 2. We assess task performance with *selection time cost* and *manipulation time cost*. The *selection time cost* measures the time participants spend selecting all target objects during the initial selection step, while the *manipulation time cost* measures the time spent manipulating all target objects to their corresponding target positions during the later manipulation step.

1.2 Results and Discussion

Before analysis, data normality is checked using Shapiro-Wilk tests and Q-Q plots. Aligned Rank Transform (ART) [4] is applied to non-normally distributed data before ANOVA. We conduct ANOVA analysis for all comparisons.

Table 1 compares the *selection time cost* of participants completing task 1 using $MOA_s(\alpha)$ with four different levels of α. Since α influences the weighting accumulation of the historical importance and current-round importance for MOA_s, it directly affects the judgment of MOA_s regarding potential target objects, leading to a change in the initial selection efficiency of MOA_s. Therefore, setting the appropriate α enhances the task performance during the initial selection in the multi-target arrangement task. A one-way repeated measures ANOVA on *selection time cost* for the four levels of α yielded a significant effect ($F_{3,45}=8.51, p=1.50\times10^{-4}, \eta_p^2=0.36$). Therefore, **H1** is supported.

Table 1: Mean ± SD values of *selection time cost* using $MOA_s(\alpha)$ with different levels of α.

α Level	*selection time cost* (s)
$\alpha=1/2$	185.4 ± 19.9
$\alpha=\mathbf{2/3}$	**165.2 ± 15.4**
$\alpha=3/4$	171.8 ± 18.1
$\alpha=1$	213.1 ± 20.5

The experimental results further show that when $\alpha=2/3$, MOA_s achieves the best task performance compared with the other three levels of α. Therefore, we set $\alpha=2/3$ to achieve optimal initial selection performance in the multi-object arrangement task.

Table 2 compares the *manipulation time cost* in task 2 using $MOA_m(\beta,\gamma)$ under different level combinations of β and γ. Since β and γ affect the visual complexity and presentation effects of Ω. Consequently, they influence the guidance efficiency of Ω in the later manipulation. Thus, different β and γ levels impact the *manipulation time cost* of MOA_m. To analyze these effects, a 5×3(β levels × γ levels) two-way repeated measures ANOVA is conducted on the *manipulation time cost* (N=16 participants, using synthetic data generated to align with means from Table 2 and readjusted standard deviations yielding moderate F-values). The analysis revealed a significant effect for

Table 2: Mean ± SD values of *manipulation time cost* (s) using $MOA_m(\beta,\gamma)$ with different levels of β and γ. Means are from the original design; SDs were readjusted (approx. 20% of mean) for the ANOVA results reported below.

	manipulation time cost (s)		
β Level	$\gamma=3$ (Thick)	$\gamma=6$ (Medium)	$\gamma=9$ (Thin)
$\beta=0$	524.8 ± 105.0	526.6 ± 105.3	522.8 ± 104.6
$\beta=6$	458.2 ± 91.6	445.4 ± 89.1	496.9 ± 99.4
$\beta=\mathbf{8}$	445.4 ± 89.1	**415.5 ± 83.1**	478.1 ± 95.6
$\beta=10$	483.7 ± 96.7	438.6 ± 87.7	464.3 ± 92.9
$\beta=+\infty$	508.6 ± 101.7	487.4 ± 97.5	473.0 ± 94.6

β ($F_{4,60}=23.12$, $p=1.35\times10^{-11}$, $\eta_p^2=0.61$), a significant effect for γ ($F_{2,30}=3.59$, $p=4.01\times10^{-2}$, $\eta_p^2=0.19$), and a significant interaction effect between β and γ ($F_{8,120}=4.89$, $p=3.06\times10^{-5}$, $\eta_p^2=0.25$). The significant interaction indicates that the effect of β on *manipulation time cost* depends on the level of γ, and vice versa. Given these significant effects, **H2** is supported.

Experimental results suggest that when $\beta=8$ and $\gamma=6$, MOA_m achieves the best task performance (415.5s). This represents an approximate 1.3× speedup compared with when the Ω guidance was disabled ($\beta=0$). While specific post-hoc comparisons would typically be performed on original experimental data to detail pairwise differences, the ANOVA results confirm the interactive influence of these coefficients.

In conclusion, $MOA_s(\alpha=2/3)+MOA_m(\beta=8,\gamma=6)$ is regarded as the optimal condition in performing multi-object arrangement tasks, as it achieves the best task performance compared with the other tested conditions in this pilot study. Therefore, in all subsequent user studies, we set MOA to $MOA_s(\alpha=2/3)+MOA_m(\beta=8,\gamma=6)$.

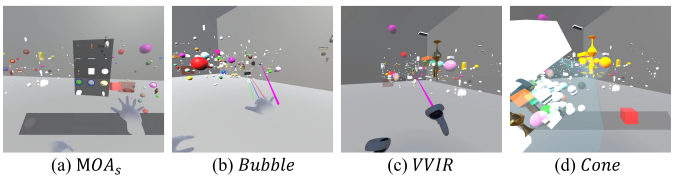
2 User Study 1: Initial Selection Evaluation

After obtaining the optimized coefficients of MOA, we conduct user study 1 to quantify the task performance and user experience of MOA's initial selection (MOA_s) in the multi-object arrangement task. Since existing multi-object initial selection methods include both controller-based and controller-free approaches, we evaluate MOA_s by comparing it with state-of-the-art methods from both controller-based and controller-free categories. We formulate a hypothesis for user study 1:

H3. Compared to state-of-the-art controller-free and controller-based multi-object initial selection methods, MOA_s achieves significant improvements in task performance, task load, and convenience during the initial selection step of the general multi-object arrangement task.

2.1 User Study Design

Participants and Apparatus. We recruit 16 participants, ten males and six females aged between 19 and 31, with normal vision or corrected-to-normal vision. Ten have experience in using HMD VR applications, and none report balance disorders. None of them are in the pilot user study. The apparatus used in this user study is the same as in the pilot user study.



(a) MOA_s (b) *Bubble* (c) *VVIR* (d) *Cone*

Fig. 2: Visualization of performing initial selection in *general scene* by using (a) the proposed MOA_s, (b) *Bubble*, (c) *VVIR*, and (d) *Cone* in *general scene*.

Condition. To comprehensively demonstrate the effectiveness of the proposed MOA_s, we compare it not only with the state-of-the-art controller-free multi-object initial selection method Bubble [5], but

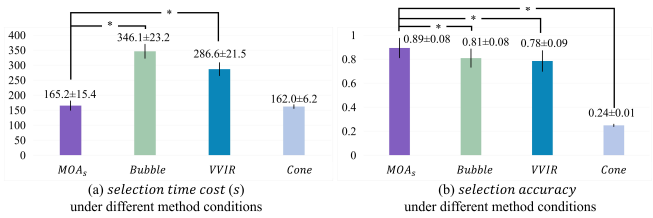


Fig. 3: Mean and standard deviation of (a) *selection time cost* and (b) *selection accuracy* in the initial selection using MOA_s, *Bubble*, *VVIR*, and *Cone*. Asterisks indicate significant differences.

also with the state-of-the-art controller-based methods VVIR [6] and Cone [1]. Therefore, the method conditions in User Study 1 include MOA_s, *Bubble*, *VVIR*, and *Cone*, as shown in Fig. 2.

Task and Procedure. The task requires participants to use all method conditions to select all target objects in *general scene*. Before starting, participants use MOA_s, *Bubble*, *VVIR*, and *Cone* to select three specified target objects from 256 candidate objects in the *general scene*, spending an average of 30 seconds on each method condition. To ensure fair comparisons, the initial positions for all conditions are set the same as those in the pilot user study. Each participant completes 4 trials. To counteract learning and fatigue effects, the presentation order of the four method conditions is counterbalanced across participants using a balanced 4×4 Latin Square design. After each trial, participants complete the NASA-TLX and SUS questionnaires. Completing all trials takes an average of 16 minutes per participant. A total of 16 (participants) × 4 (method conditions) = 64 trials are collected.

Metrics. We use the objective metrics *selection time cost* and *selection accuracy* to quantify the task performance of the initial selection step in the multi-object arrangement task. The *selection time cost* is defined in the metrics of Section 1.1. The *selection accuracy* refers to the ratio of the number of target objects selected by the participants to the total number of objects selected. To evaluate task load, we use the standard NASA-TLX questionnaire [7]. To evaluate convenience, we use the System Usability Scale (SUS) [8] for each method condition.

2.2 Results and Discussion

We examine the normal distribution of the data using Shapiro-Wilk tests and Q-Q plots before analysis, and utilize the ART to perform transformations for non-normally distributed data. Then, we conduct ANOVA analyses for all comparisons, reporting effect sizes wherever feasible. Additionally, we perform Bonferroni post-hoc analyses to examine individual differences between MOA_s and (*Bubble*, *VVIR*, *Cone*).

Table 3: Post-hoc analysis of between MOA_s and other conditions for task performance metrics in user study 1 using Bonferroni.

metric	comparison		mean dif.	std. dif.	p-value
selection time cost	MOA_s	*Bubble*	-178.8	6.2	6.0×10^{-36}
		VVIR	-119.4	6.2	1.9×10^{-26}
		Cone	5.3	6.2	1.0
selection accuracy	MOA_s	*Bubble*	0.8	0.0	7.0×10^{-3}
		VVIR	0.1	0.0	3.0×10^{-4}
		Cone	0.2	0.0	1.6×10^{-23}

Task Performance. Regarding task performance, Fig. 3 visualizes the comparisons of *selection time cost* and *selection accuracy* during the initial selection step of the multi-object arrangement task in *general scene* under four different conditions. The effect test across the four conditions for *selection time cost* yields ($F_{3,45} = 425.45$, $p = 2.26 \times 10^{-40}$, $\eta_p^2 = 0.95$), and for *selection accuracy* yields ($F_{3,45} = 279.79$, $p = 3.22 \times 10^{-35}$, $\eta_p^2 = 0.93$), indicating significant differences among the conditions in task performance. Table 3 shows the post-hoc statistical results comparing MOA_s with the other three conditions for both *selection time cost* and *selection accuracy*, using the Bonferroni method. MOA_s quickly extracts potential target objects in highly occluded scenes due to its object importance-driven selection mechanism and reduces interference and uncertainty during the initial selection step thanks to an efficient interaction mode that enables simultaneous selection. Compared with *Bubble* and *VVIR*, MOA_s achieves significant improvements in both *selection time cost* and *selection accuracy*, demonstrating clear advantages in task performance over these conditions.

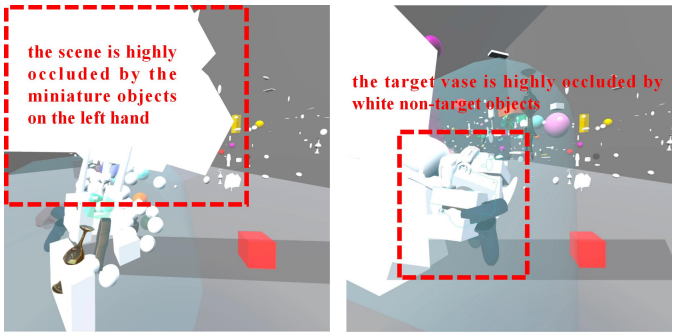


Fig. 4: Visualization of performing initial selection using *Cone* in *general scene*.

Table 4: Statistical results of scores (mean value ± standard deviation) in NASA-TLX questionnaire under different conditions in user study 1.

QID	Mean ± SD NASA-TLX scores			
	MOA_s	*Bubble*	*VVIR*	*Cone*
Q1	28.1 ± 5.7	30.6 ± 6.6	28.9 ± 8.5	43.8 ± 8.9
Q2	58.1 ± 12.4	60.3 ± 11.0	58.1 ± 12.4	57.5 ± 11.8
Q3	58.8 ± 5.0	70.0 ± 10.7	68.1 ± 9.8	63.1 ± 7.7
Q4	30.3 ± 7.4	47.5 ± 4.8	35.6 ± 5.4	55.6 ± 11.8
Q5	48.1 ± 11.7	54.7 ± 7.6	52.5 ± 8.4	58.4 ± 8.1
Q6	32.2 ± 11.8	47.2 ± 10.0	53.1± 11.2	62.2 ± 10.6
TOTAL	42.6 ± 6.1	51.7 ± 4.3	49.1 ± 4.3	56.7 ± 3.6

Fig. 4 illustrates the initial selection process using *Cone* in a *general scene*. The scale of the selectable region is set by the default farthest function [1]. Due to the dense distribution of objects in the *general scene* and the wide variation in object sizes, target objects are often heavily occluded by non-target objects (marked in white) within the minimal 3D capsule for selection, as shown in Fig. 4 (a). When a participant attempts to select the target vase within this minimal capsule using the controller, the extensive overlap causes multiple non-target objects to be selected alongside the target vase, resulting in a high number of erroneous selections, as depicted in Fig. 4 (b). As a result, although the *selection time cost* for *Cone* does not differ significantly from that of MOA_s, the *selection accuracy* of *Cone* is 3.7× lower than that of MOA_s. With a selection accuracy below 30%, *Cone* produces excessive invalid selections, limiting its effective application. Therefore, MOA_s demonstrates a significant advantage over *Cone* in terms of task performance. Thus, we obtain **Conclusion 1**: MOA_s significantly improves the task performance with all state-of-the-art controller-free and controller-based multi-object selection methods in the initial selection step of the multi-object arrangement task in *general scene*.

Task Load. Regarding task load, Table 4 details the NASA-TLX questionnaire scores under different conditions. The effect test across the four conditions for the NASA-TLX total score yields ($F_{3,45} = 25.77$, $p = 7.76 \times 10^{-11}$, $\eta_p^2 = 0.58$), indicating significant differences in task load among conditions.

Table 5 presents the post-hoc statistical results comparing MOA_s with the other three conditions for the NASA-TLX score. Based on participants' report, although MOA_s requires the additional activation

Table 5: Post-hoc analysis of between MOA_s and other conditions for the NASA-TLX total score in user study 1 using Bonferroni.

metric	comparison		mean dif.	std. dif.	p-value
Q1	MOA_s	*Bubble*	-2.5	2.7	9.6×10^{-1}
		VVIR	-0.8		1.0
		Cone	-15.7		7.5×10^{-16}
Q2	MOA_s	*Bubble*	-2.2	4.1	1.0
		VVIR	0.0		1.0
		Cone	0.6		1.0
Q3	MOA_s	*Bubble*	-11.2	3.0	3.0×10^{-2}
		VVIR	-9.3		1.8×10^{-2}
		Cone	-4.3		9.3×10^{-1}
Q4	MOA_s	*Bubble*	-17.2	2.8	3.7×10^{-7}
		VVIR	-5.3		3.7×10^{-1}
		Cone	-25.3		4.0×10^{-12}
Q5	MOA_s	*Bubble*	-6.6	3.2	2.7×10^{-1}
		VVIR	-4.4		9.3×10^{-1}
		Cone	-10.3		1.3×10^{-2}
Q6	MOA_s	*Bubble*	-15.0	3.5	3.4×10^{-4}
		VVIR	-20.9		6.0×10^{-7}
		Cone	-30.0		2.1×10^{-11}
TOTAL	MOA_s	*Bubble*	-9.1	1.6	4.0×10^{-6}
		VVIR	-6.5		1.0×10^{-2}
		Cone	-14.1		2.5×10^{-11}

of the candidate panel to conveniently select objects, partially offsetting its advantage in physical demands, its score in Q2 (physical demands) is comparable to that of the state-of-the-art controller-free and controller-based methods. Although *Bubble* supports parallel multi-object selection, its reliance on simultaneous multi-finger pointing interactions makes it highly prone to selection errors in target-dense and occluded scenes. MOA_s effectively circumvents selection difficulties in highly occluded environments by efficiently rearranging target objects onto a candidate panel. Despite the initial step of generating the panel, MOA_s performs comparably to *Bubble* in terms of mental demand (Q1) and physical demand (Q5). However, MOA_s eliminates the repeated attempts and error corrections caused by occlusion, leading to significant advantages in terms of time pressure (Q3), self-perceived performance (Q4), and frustration (Q6). Ultimately, these advantages collectively contribute to MOA_s significantly outperforming *Bubble* in NASA-TLX total score.

In comparison to the controller-based method condition *VVIR*, according to participants' feedback, *VVIR* exhibits high efficiency in selecting a small number of multiple targets due to its ability to select targets with small controller movements. This precision of single-point operation allows MOA_s to perform comparably to *VVIR* in terms of mental demand (Q1), self-perceived performance (Q4), and physical demand (Q5). However, when the task extends to a large number of multiple targets, *VVIR* requires participants to make repeated and fine-grained aiming adjustments to select them one by one. In contrast, MOA_s' parallel selection mode allows participants to naturally and efficiently confirm multiple targets at once, resulting in overwhelmingly significant advantages in terms of time pressure (Q3) and frustration (Q6). Consequently, this makes MOA_s significantly better than *VVIR* in NASA-TLX total score.

In comparison to another controller-based method condition *Cone*, *Cone* employs a strategy of "broadly selecting with a cylinder first, and then finely picking from within." This macro-capture mechanism provides time efficiency when dealing with spatially concentrated target groups, thus MOA_s performs comparably to *Cone* in terms of time pressure (Q3). However, MOA_s allows participants to directly and precisely click on targets in the generated panel, avoiding the cumbersome two-step operation of "coarse selection followed by fine selection" inherent in *Cone*. This fundamental simplification of the interaction flow greatly reduces the participant's cognitive load and operational uncertainty, resulting in significant advantages for MOA_s in terms of mental demand (Q1), self-perceived performance (Q4), physical demand (Q5), and frustration (Q6). Ultimately, these comprehensive experience improvements make MOA_s significantly better than *Cone* in NASA-TLX total score.

The p-values indicate that MOA_s achieves a significantly lower NASA-TLX total score than all other conditions. Therefore, we conclude **Conclusion 2**: Compared with all state-of-the-art controller-free and controller-based multi-object selection methods, MOA_s significantly reduces task load in the initial selection step of the general multi-object arrangement task.

Table 6: Statistical results of scores (mean value $\pm$ standard deviation) in SUS under different conditions in user study 1.

QID	Mean $\pm$ SD SUS scores			
	MOA_s	*Bubble*	*VVIR*	*Cone*
Q1	3.7 ± 0.9	3.2 ± 1.0	3.3 ± 0.7	3.3 ± 0.7
Q2	1.9 ± 0.7	2.4 ± 0.6	2.2 ± 0.7	2.4 ± 0.6
Q3	3.3 ± 1.0	3.3 ± 0.9	3.4 ± 0.9	2.9 ± 0.3
Q4	2.2 ± 0.7	2.7 ± 0.7	2.3 ± 0.7	3.0 ± 0.4
Q5	3.7 ± 0.8	3.3 ± 0.8	3.5 ± 0.7	3.0 ± 0.5
Q6	1.1 ± 0.3	1.4 ± 0.6	1.6 ± 0.6	1.7 ± 0.5
Q7	4.4 ± 0.7	3.6 ± 0.9	3.5 ± 0.6	3.2 ± 0.8
Q8	1.1 ± 0.3	1.3 ± 0.4	1.2 ± 0.4	2.4 ± 0.7
Q9	4.2 ± 0.8	3.1 ± 0.8	3.3 ± 0.7	3.0 ± 1.0
Q10	1.0 ± 0.1	1.3 ± 0.7	1.1 ± 0.5	2.6 ± 0.5
TOTAL	79.2 ± 5.7	68.8 ± 6.5	71.6 ± 5.2	57.8 ± 5.0

Table 7: Post-hoc analysis of between MOA_s and other conditions for the SUS total score in user study 1 using Bonferroni.

metric	comparison		mean dif.	std. dif.	p-value
SUS	MOA_s	*Bubble*	10.5	0.7	5.1×10^{-6}
		VVIR	7.8	0.6	1.5×10^{-4}
		Cone	12.2	0.1	3.4×10^{-8}

Convenience. In terms of convenience, Table 6 details the SUS statistical results under different method conditions. The effect test across the four conditions for SUS yields ($F_{3,45} = 16.20$, $p = 7.92 \times 10^{-8}$, $\eta_p^2 = 0.45$), indicating significant differences in convenience among the conditions. Compared to other method conditions, MOA_s achieves a higher SUS total score. Participant feedback demonstrates that MOA_s effectively presents target objects on the panel for participants to select in the complex scene with high density and occlusion. This instills a strong sense of confidence in participants, leading them to perceive the system not only as easy to use but also as enabling them to efficiently complete a large number of target object selection tasks in complex scenes, resulting in highly positive SUS total scores. Table 7 presents the post-hoc statistical results comparing MOA_s with the other three conditions for the SUS total score. The p-values indicate that MOA_s achieves a significantly SUS total score than all other methods. Therefore, we conclude **Conclusion 3**: Compared with all state-of-the-art controller-free and controller-based multi-object selection methods, MOA_s significantly improves convenience in the initial selection step of the general multi-object arrangement task. Thus, based on **Conclusion 1**, **Conclusion 2**, and **Conclusion 3**, **H3** is supported.

3 User Study 2: Later Manipulation Evaluation

In this section, we conduct user study 2 to further evaluate the task performance, task load, and convenience of MOA's later manipulation step (MOA_m) in the multi-object arrangement task, and to compare

MOA_m with state-of-the-art controller-free and controller-based later manipulation methods. We formulate the hypothesis for user study 2 as follows:

H4. Compared to state-of-the-art later manipulation methods, MOA_m significantly improves task performance, reduces task load, and enhances convenience in the later manipulation step of the general multi-object arrangement task.

3.1 User Study Design

Participants and Apparatus. The same 16 participants from user study 1 are also recruited for user study 2, and the system setup remains consistent between the two user studies.

Condition. To comprehensively demonstrate the effectiveness of the proposed MOA_m, we compare it not only with an existing controller-free method but also with a state-of-the-art controller-based later manipulation method. We use the intuitive Object Proxy method as the controller-free comparison, and VVIR as the controller-based comparison. Thus, the method conditions in user study 2 include MOA_m, *Object Proxy*, and *VVIR*, as shown in Fig. 6.

Task and Procedure. The task requires participants to manipulate all target objects to designated target positions in *general scene* using each of the method conditions. Before the task begins, participants spend 1.5 minutes practicing with MOA_m, *Object Proxy*, and *VVIR* to manipulate three selected target objects to their corresponding target positions in the *general scene*. To ensure fair comparisons, the initial object positions for all conditions are identical to those used in previous user studies. Each participant completes 3 trials. The order of these three conditions is counterbalanced across participants using all six possible permutations ($3! = 6$ orders). Participants are systematically assigned to one of these sequences to mitigate order effects. After each trial, participants complete the NASA-TLX and SUS questionnaires. The entire set of trials takes approximately 25 minutes per participant. A total of 16 (participants) $\times$ 3 (method conditions) = 48 trials are collected.

Metrics. We use the objective metric *manipulation time cost* to quantify task performance in the later manipulation step of the multi-object arrangement. The definition of *manipulation time cost* is provided in Section 1. As in user study 1, we employ the standard NASA-TLX questionnaire and SUS to evaluate task load and convenience. Details on the NASA-TLX questionnaire and SUS can be found in Section 2.1.

3.2 Results and Discussion

We first assess the normality of the data using Shapiro-Wilk tests and Q-Q plots. For non-normally distributed data, we apply the ART to facilitate transformations. Subsequently, we conduct ANOVA analyses for all comparisons and report effect sizes where applicable. We also perform Bonferroni post-hoc analyses to evaluate individual differences between MOA_m and the other conditions (*Object Proxy*, *VVIR*).

Task Performance. In terms of task performance, Fig. 5 visualizes the *manipulation time cost* comparisons in the later manipulation step of the multi-object arrangement task in the *general scene* under three method conditions. The effect test for three conditions in *manipulation time cost* yields ($F_{2,30} = 167.25$, $p = 1.46 \times 10^{-21}$, $\eta_p^2 = 0.88$), indicating significant differences among the conditions in *manipulation time cost*. Thanks to MOA_m's ability to provide corresponding coarse and fine manipulation modes based on different manipulation contexts, participants can quickly manipulate target objects close to the desired target position using the coarse manipulation phase, while the fine manipulation phase allows precise positioning of the object from near the target location to the final target position.

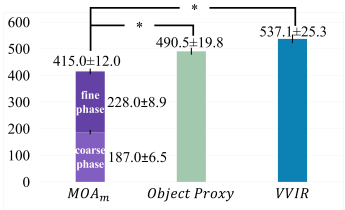
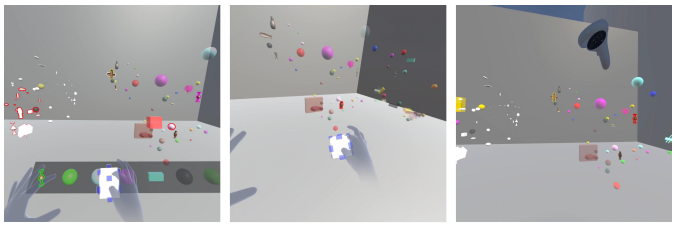


Fig. 5: Mean and standard deviation of *manipulation time cost* in later manipulation using MOA_m, *Object Proxy*, and *VVIR*.

Although fine phase involves less operational distances and angles, its requirement for precision leads to significant time expenditure. Fig. 5 shows that fine manipulation accounts for 80% of the total *manipulation time cost*. However, relying solely on the fine phase makes it difficult to quickly manipulate the object to the vicinity of the target location. Dividing later manipulation into coarse and fine modes is necessary. According to participants feedback, the two-phase manipulation mode of MOA_m make the manipulation process more flexible and efficient, with each mode enabling the realization of specific goals based on the participants' intentions.



(a) MOA_m (b) *Object Proxy* (c) *VVIR*

Fig. 6: Visualization of performing later manipulation in *general scene* by using (a) the proposed MOA_m, (b) *Object Proxy*, and (c) *VVIR*.

Table 8: Post-hoc analysis of between MOA_m and other conditions for the task performance metric in user study 2 using Bonferroni.

metric	comparison		mean dif.	std. dif.	p-value
manipulation time cost	MOA_m	*Object Proxy*	-101.3	7.1	9.6×10^{-18}
		VVIR	-122.1	7.1	8.0×10^{-21}

The post-hoc statistical results in Table 8 show that MOA_m achieves a significant improvement in *manipulation time cost* compared to state-of-the-art method conditions. Therefore, we establish **Conclusion 4**: MOA_m significantly improves task performance compared to state-of-the-art controller-free and controller-based methods in the later manipulation step of the general multi-object arrangement task.

Table 9: Mean $\pm$ SD scores of each question in NASA-TLX questionnaire under different conditions in user study 2.

QID	Mean $\pm$ SD NASA-TLX scores		
	MOA_m	*Object Proxy*	*VVIR*
Q1	27.8 $\pm$ 5.2	30.6 $\pm$ 5.1	31.3 $\pm$ 5.9
Q2	60.0 $\pm$ 8.9	70.0 $\pm$ 8.0	74.4 $\pm$ 9.6
Q3	65.6 $\pm$ 6.0	68.8 $\pm$ 9.2	80.0 $\pm$ 8.6
Q4	30.3 $\pm$ 5.3	33.1 $\pm$ 4.0	57.2 $\pm$ 8.4
Q5	45.6 $\pm$ 4.4	50.3 $\pm$ 8.1	49.4 $\pm$ 8.9
Q6	34.4 $\pm$ 6.8	40.6 $\pm$ 4.4	46.9 $\pm$ 6.4
TOTAL	44.0 $\pm$ 3.3	48.9$\pm$ 3.4	56.5 $\pm$ 4.0

Task Load. Regarding task load, Table 9 presents the detailed NASA-TLX questionnaire results from user study 2. The effect test for three conditions on the NASA-TLX total score yields ($F_{2,30} = 48.71$, $p = 5.52 \times 10^{-12}$, $\eta_p^2 = 0.68$), indicating significant differences among conditions in task load.

Table 10 shows the post-hoc statistical results comparing MOA_m with the other two conditions for the NASA-TLX score. According to participants' feedback, compared with *Object Proxy* that requires directly manipulating a large number of target objects, MOA_m can efficiently manipulate target objects to the vicinity of the corresponding target positions, benefiting from the built-in auxiliary structure. This leaves the precise tuning of complex multi-object manipulation to a more sensitive fine phase of MOA_m, significantly reducing the physical and cognitive burden on participants. Consequently, MOA_m

Table 10: Post-hoc analysis of between MOA_m and other conditions for the NASA-TLX total score in user study 2 using Bonferroni.

metric	comparison		mean dif.	std. dif.	p-value
Q1	MOA_m	*Object Proxy*	-2.8	1.9	4.5×10^{-1}
		VVIR	-3.5		2.4×10^{-1}
Q2	MOA_m	*Object Proxy*	-10.0	3.1	8.0×10^{-4}
		VVIR	-14.4		1.2×10^{-4}
Q3	MOA_m	*Object Proxy*	-3.2	2.8	8.4×10^{-1}
		VVIR	-14.4		2.3×10^{-5}
Q4	MOA_m	*Object Proxy*	-2.8	2.2	6.1×10^{-1}
		VVIR	-26.9		1.6×10^{-15}
Q5	MOA_m	*Object Proxy*	-4.7	2.6	2.4×10^{-1}
		VVIR	-3.8		4.8×10^{-1}
Q6	MOA_m	*Object Proxy*	-6.3	2.1	1.4×10^{-2}
		VVIR	-12.5		1.0×10^{-6}
TOTAL	MOA_m	*Object Proxy*	-4.9	1.3	1.0×10^{-3}
		VVIR	-12.5		2.9×10^{-12}

scores better than *Object Proxy* on all six dimensions (Q1-Q6) of the NASA-TLX, especially in Q2 (Physical Demands) and Q6 (Frustration), achieving significant advantages of 14.3% and 15.3%, respectively.

VVIR integrates a single-step operation mode that forces participants to repeat the entire "aim, move, and place" cycle for each target object when dealing with a large number of target objects. This high frequency of repetitive labor not only leads to participant fatigue but also reduces manipulation accuracy, ultimately making participants feel that they cannot control the entire interaction process. MOA_m's two-stage design effectively decouples the complex multi-object manipulation task into coarse and fine phases. First, participants can use guiding structures to quickly and batch manipulate multiple target objects to the vicinity of their respective target locations. Subsequently, they can focus on subsequent fine-grained adjustments. This strategy improves the efficiency and smoothness of the later manipulation task. Therefore, MOA_m shows improvement over *VVIR* in all question items of NASA-TLX, particularly in Physical Demands (Q2), Temporal Demands (Q3), Own Performance (Q4), and Frustrations (Q6), with gains of 19.4%, 18.0%, 47.0%, and 26.7%, respectively. The p-values indicate that MOA_m achieves significantly better NASA-TLX scores than *Object Proxy* and *VVIR*. Therefore, we obtain **Conclusion 5**: Compared to state-of-the-art controller-free and controller-based methods, MOA_m significantly reduces task load in the later manipulation step of the general multi-object arrangement task.

Convenience. In terms of convenience, Table 11 presents detailed SUS statistical results. The effect test for three conditions on SUS scores yields ($F_{2,30} = 14.14$, $p = 1.72 \times 10^{-5}$, $\eta_p^2 = 0.39$), indicating significant differences among conditions in convenience.

According to participant feedback, MOA_m facilitates the later manipulation step through a simple pinch-and-release gesture. Furthermore, guided by the auxiliary structure, MOA_m enables participants to precisely manipulate multiple objects to their designated target positions. Participants perceive MOA_m is more effective than *Object Proxy* and *VVIR* in multi-object later manipulation tasks. As a result, MOA_m outperforms the other two method conditions in the SUS total score. Table 12 shows the post-hoc statistical results comparing MOA_m with the other two conditions for the SUS total score. The p-values indicate that MOA_m achieves significantly higher SUS scores than both *Object Proxy* and *VVIR*. Therefore, we establish **Conclusion 6**: Compared with state-of-the-art controller-free and controller-based methods, MOA_m significantly improves convenience in the later manipulation step of the general multi-object arrangement task. Thus, based on **Conclusion 4**, **Conclusion 5**, and **Conclusion 6**, the results support **H4**.

Table 11: Mean ± SD scores of each question in SUS under different conditions in user study 2.

QID	Mean ± SD SUS scores		
	MOA_m	*Object Proxy*	*VVIR*
Q1	3.9 ± 0.6	3.4 ± 1.0	3.4 ± 0.8
Q2	1.8 ± 0.7	1.9 ± 0.57	2.2 ± 0.7
Q3	3.4 ± 0.9	3.3 ± 0.8	3.3 ± 0.9
Q4	1.8 ± 0.6	2.3 ± 0.7	2.9 ± 0.9
Q5	3.8 ± 0.7	3.6 ± 0.5	3.5 ± 0.63
Q6	1.1 ± 0.3	1.3 ± 0.5	1.4 ± 0.5
Q7	4.3 ± 0.7	3.8 ± 0.8	3.8 ± 0.7
Q8	1.3 ± 0.5	1.4 ± 0.6	1.6 ± 0.8
Q9	4.3 ± 0.7	3.8 ± 0.7	3.4 ± 0.6
Q10	1.3 ± 0.6	1.3 ± 0.5	1.1 ± 0.3
TOTAL	81.4 ± 5.2	73.9 ± 6.4	70.8 ± 6.0

Table 12: Post-hoc analysis of between MOA_m and other conditions for the the SUS total score in user study 2 using Bonferroni.

metric	comparison		mean dif.	std. dif.	p-value
SUS	MOA_m	*Object Proxy*	7.5	2.1	2.0×10^{-3}
		VVIR	10.8	2.1	1.5×10^{-5}

REFERENCES

[1] M. Maslych, Y. Hmaiti, R. Ghamandi, P. Leber, R. K. Kattoju, J. Belga, and J. J. LaViola, "Toward intuitive acquisition of occluded vr objects through an interactive disocclusion mini-map," in *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2023, pp. 460–470. 1, 3

[2] J. Bergström, T.-S. Dalsgaard, J. Alexander, and K. Hornbæk, "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*, 2021, pp. 1–20. 1, 2

[3] M. R. T. Grady Andersen, "The ultimate guide to vr user testing-bestpractices and essential tools," 2025. [Online]. Available: https://moldstud.com/articles/p-the-ultimate-guide-to-vr-user-testing-best-practices-and-essential-tools 1

[4] F. Welsford-Ackroyd, A. Chalmers, R. Kuffner dos Anjos, D. Medeiros, H. Kim, and T. Rhee, "Spectator view: Enabling asymmetric interaction between hmd wearers and spectators with a large display," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. ISS, pp. 1–17, 2021. 2

[5] W. Delamare, M. Daniel, and K. Hasan, "Multifingerbubble: A 3d bubble cursor variation for dense environments," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 2022, pp. 1–6. 2

[6] Q. Zheng, L. Wang, W. Ke, and S. K. Im, "Vvir-om: Efficient object manipulation in vr with variable virtual interaction region," *International Journal of Human–Computer Interaction*, pp. 1–14, 2023. 3

[7] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183. 3

[8] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996. 3