Exploration of Bare-Hand Mid-Air Pointing Selection Techniques for Dense Virtual Reality Environments

Rongkai Shi*
Xi'an Jiaotong-Liverpool Univeristy
Suzhou, China
rongkai.shi19@student.xjtlu.edu.cn

Jialin Zhang*
Xi'an Jiaotong-Liverpool Univeristy
Suzhou, China
jialin.zhang21@student.xjtlu.edu.cn

Yong Yue
Xi'an Jiaotong-Liverpool Univeristy
Suzhou, China
yong.yue@xjtlu.edu.cn

Lingyun Yu
Xi'an Jiaotong-Liverpool Univeristy
Suzhou, China
lingyun.yu@xjtlu.edu.cn

Hai-Ning Liang†
Xi'an Jiaotong-Liverpool Univeristy
Suzhou, China
haining.liang@xjtlu.edu.cn

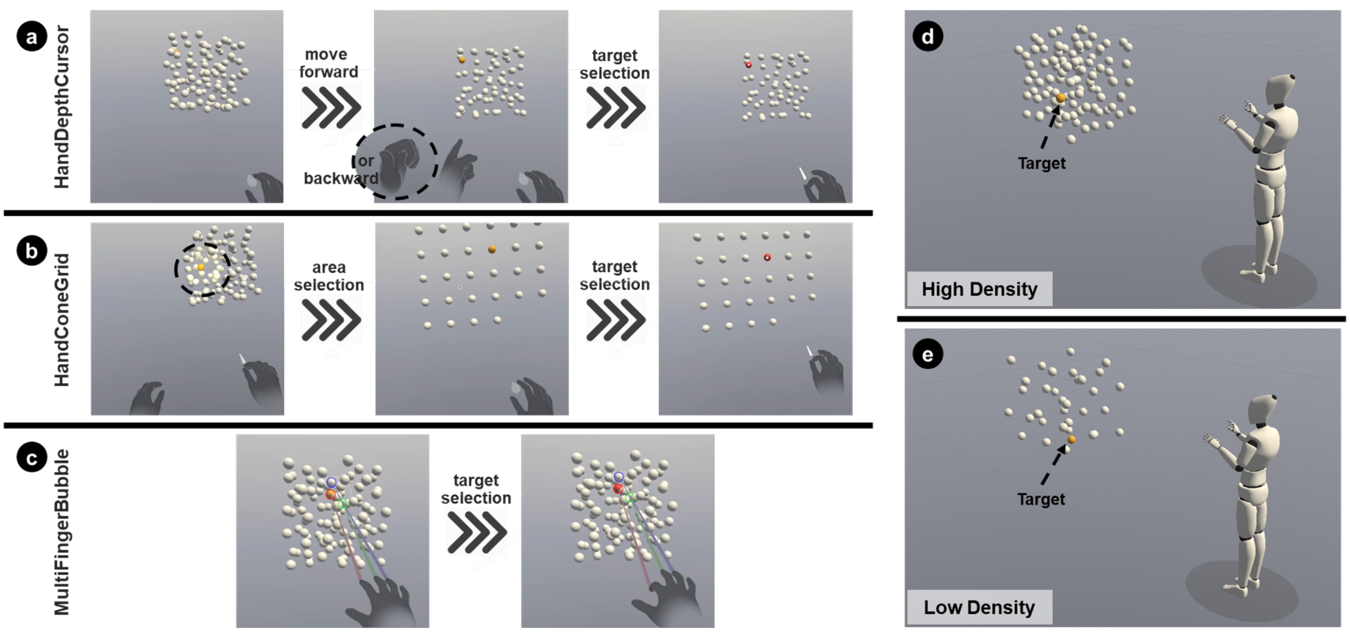

Figure 1: Screenshots of the two new bare-hand mid-air pointing/selection techniques for dense VR environments: (a) *HandDepthCursor*, and (b) *HandConeGrid*. We compare them with (c) *MultiFingerBubble* [9], an existing hand gesture-based technique, in a controlled target selection task with (d-e) two levels of density. Note: black lines, arrows, texts, and avatars are for illustrations only and are not shown to users.

ABSTRACT

Target selection in dense virtual reality (VR) environments is challenging. Prior work has explored different controller-based raycasting techniques to assist target selection in such environments. However, limited research has focused on selection via mid-air barehand, which represents another major input metaphor for immersive environments. In this paper, we first review the existing raycasting selection techniques for dense VR environments. Based on this, we propose and develop two freehand pointing selection techniques—HandDepthCursor and HandConeGrid, and implement MultiFingerBubble, a recently-proposed technique. We then conduct a user study to compare and evaluate their performance and experience in a target selection task in dense VR environments. Our results suggest that HandDepthCursor and HandConeGrid led to significantly faster and more accurate selection performance, and lower perceived workload and arm fatigue. In addition, HandConeGrid showed a distinct advantage in high-density environments.

*Both authors contributed equally to this research.
†Corresponding author (haining.liang@xjtlu.edu.cn)

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585615

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; **Empirical studies in interaction design**; **Interaction techniques**.

KEYWORDS

virtual reality, object selection, pointing selection, dense environment, bare-hand interaction, head-mounted displays

ACM Reference Format:
Rongkai Shi, Jialin Zhang, Yong Yue, Lingyun Yu, and Hai-Ning Liang. 2023. Exploration of Bare-Hand Mid-Air Pointing Selection Techniques for Dense Virtual Reality Environments. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3585615

1 INTRODUCTION

Object selection is a canonical task in immersive virtual reality (VR) and other immersive environments and commonly serves as an initial step for various interactions [1, 3, 14, 19]. Pointing based on raycasting [21] is a major object selection metaphor in such environments [1, 14]. When using a raycasting-based technique, users control a ray emitted from an origin that is commonly from controllers [14, 35, 37] but can also be from users' hands [9, 24, 36, 41], head [20, 25, 33, 39, 40], or eyes [17, 22, 24, 25, 33]. Raycasting-based techniques allow users to select objects beyond their reach with relatively less physical movement [1]. Due to these advantages, raycasting is widely used in current commercial VR head-mounted displays (HMDs). However, raycasting's performance deteriorates in dense environments due to two types of ambiguities: (1) the objects might be small and distant and positioned close to each other, making the target challenging to be located, and (2) the target may be partially or even fully occluded by other objects [32, 38, 42].

To overcome these challenges in dense environments, researchers have proposed enhancement techniques to support pointing selection (e.g., [11, 13, 42]; see also Section 2). For example, Kopper et al. [13] proposed Sphere-casting refined by QUAD-menu (SQUAD) to let users refine the selection progressively until they select the intended target. Like SQUAD, most enhancement techniques are designed for and evaluated with handheld controllers, which are currently the most common input for interacting with VR HMDs [19]. There has been less research that has looked into bare-hand mid-air input for VR, an alternative to controllers. With advances in hand tracking technologies, bare-hand mid-air input has become more accurate and cost-effective [6, 15]. Some state-of-the-art VR HMDs, such as Meta Quest 2 and PICO 4, have inside-out cameras enabling bare-hand mid-air input without extra hand-tracking devices. Such a controller-free approach is becoming more popular because it may facilitate effective interaction with increased immersion and presence [6]. It is more similar to how we interact with real-world objects and increases expressiveness. However, gesture-based interaction is different from controller-based interactions—6 degrees of freedom (DoF; 3 positional DoF and 3 rotational DoF) with extended DoF and operational possibilities when including buttons and triggers [15, 19]. This leads to a gap in bare-hand pointing selection for dense VR environments.

To address this need, we first review existing controller-based pointing techniques and classify them into a taxonomy [1, 42]. Based on this survey, we developed two potential bare-hand pointing selection techniques, HandDepthCursor and HandConeGrid, with depth- and grid-based disambiguation mechanisms achieved via expressive hand gestures. Moreover, we also implemented MultiFingerBubble [9], a recently-proposed technique, and evaluated them via a user study ($N = 18$) with two levels of density. Measurements include efficiency, accuracy, perceived user experience, workload, arm fatigue, and participants' preference. Our results find that HandDepthCursor and HandConeGrid outperformed MultiFingerBubble in accuracy, perceived workload, and arm fatigue. HandConeGrid showed significantly higher accuracy in high-density environments. MultiFingerBubble, though faster than the other two, generated more errors and required more effort.

The main contribution of this work is an empirical evaluation of three bare-hand mid-air pointing techniques for dense environments on VR HMDs. Our work provides insights into the design and development of future freehand selection techniques for VR HMDs.

2 RELATED WORK

We review the enhanced raycasting techniques based on the taxonomy proposed by Argelaguet and Andujar [1], which divides the disambiguation mechanisms into three groups: *manual, heuristic*, and *behavioral*.

Manual approaches involve additional steps to disambiguate the context, making the target obvious to observe and easier to select. The existing techniques can be further divided into *depth-based* and *grid-based* approaches [42]. A depth-based approach adds a movable depth cursor to the ray, which enables the selection on the depth axis [11, 32]. Baloup et al. [2] proposed RayCursor for VR HMDs through which users can move the cursor via the controller's touchpad. Yu et al. [42] introduced Alpha Cursor, an enhanced version of RayCursor, and when using it, the objects between the cursor and the user are made transparent to help with fully-occluded object selection. On the other hand, a grid-based approach enables users to select an area where the intended target is located, then rearranges the objects within the area on a grid. The main difference among the existing techniques is in the area selection step. SQUAD [13] uses a quad-menu to provide progressive refinement. Expand [7] allows users to zoom in on a specific area. Flower Cone [42] extends the ray to an adjustable cone (or a spotlight) for the area selection. Lasso Grid [42] defines the region by drawing a shape. Besides, Flower Ray [11] lists all the intersected objects by the ray on the grid, while Grid Wall [42] lists all objects in the scene on the grid. For techniques that follow manual approaches, different elements of the handheld controller are mapped to these functionalities. For example, the joystick is commonly used for adjusting the area's size or the depth cursor, while buttons/triggers can activate discrete and continuous selection mechanisms by short-pressing or long-pressing them, respectively.

Heuristics approaches rank the objects based on defined heuristics and determine the target that the user intends to select. Flashlight [16] selects the object that is closest to the central axis of the selection volume. Sticky-Ray [30] remains the last intersected

Table 1: Summary of the literature review: existing controller-based and our bare-hand raycasting techniques for dense VR environments.

	Disambiguation Mechanism	Key Literature
Controller	Manual (Grid-based)	SQUAD [13], Expand [7], Flower Ray [11], Flower Cone [42], Grid Wall [42], Lasso Grid [42]
	Manual (Depth-based)	Depth ray [11, 32], Lock ray [11], RayCursor [2], Alpha Cursor [42]
	Heuristic	Flashlight [16], Sticky-Ray [30], Bubble Ray [18]
	Behavioral	IntenSelect [8], SenseShape [23]
Bare-Hand	Manual (Grid-based)	HandConeGrid (ours)
	Manual (Depth-based)	HandDepthCursor (ours)
	Behavioral	MultiFingerBubble [9]

object as the candidate object. Lu et al. [18] proposed Bubble Ray, a raycasting technique that selects the nearest target in the spherical bubble. Besides, assigning probabilities when performing selection is another approach [27, 40]. Instead of applying heuristics for the selection confirmation, behavioral approaches consider users' actions when manipulating the ray to locate the target. IntenSelect [8] and SenseShape [23] use time-based ranking approaches to calculate users' intentions dynamically. A notable behavioral technique is MultiFingerBubble [9], the only bare-hand pointing technique we found for target selection in dense VR environments. MultiFingerBubble maps several objects in the selection volume to the user's fingers. By moving the hand and fingers, the user can adjust the selection volume and the candidate objects. Once the target is determined, the user can flex the corresponding finger to confirm the selection.

Table 1 summarizes the above techniques designed for selection in dense VR environments. Prior work has shown that some manual approaches show greater efficiency than heuristics and behavioral approaches due to the reduced time requirements for disambiguation [11]. For this work, we developed two bare-hand mid-air pointing techniques based on the manual approach to allow us to do a first evaluation of both depth-based and grid-based sub-approaches. In addition, we also implemented the first and only behavioral-based technique, MultiFingerBubble [9], and conducted a user study to compare and evaluate their performance and user experience.

3 TECHNIQUES

3.1 HandDepthCursor (HDC)

HandDepthCursor (HDC) follows a depth-based manual approach and was inspired by RayCursor [2] and Alpha Cursor [42]. Figure 1a shows a preview of HDC. HDC uses raycasting as the selection mechanism. The user can point to the target using her dominant hand, and perform selection via a pinch gesture. Like other depth-based techniques, the user can go deeper and move back to the densely cluttered environment via two non-dominant hand gestures. To go deeper, the user can point her index finger to the front. To come close, the user can point her thumb to the back. These gestures are similar to their expressions in the real world. Besides, we also let the objects become transparent if they are behind the cursor, which can further improve the disambiguation of objects in occluded scenarios [42].

3.2 HandConeGrid (HCG)

HandConeGrid (HCG, as shown in Figure 1b) follows a grid-based approach using a cone for area selection like Flower Cone [42]. HCG starts with cone-casting. The user uses her non-dominant hand to cast a cone to cover a volume that the target locates. By spreading or closing her non-dominant hand (more specifically, spreading fingers or putting fingertips together), the user can adjust the size of the volume. To confirm the area selection, the user performs a pinch gesture with her dominant hand. All objects in the defined volume are listed on a grid. Then, the user enters the target selection step where she points to the target and performs another pinch gesture to confirm the selection. If the user wants to redo an area selection, she can pinch when pointing to a blank area on either side of the grid.

3.3 MultiFingerBubble (MFB)

MultiFingerBubble (MFB) was adopted from [9] that follows a behavioral approach (see Figure 1c). With MFB, a user casts a spherical volume positioned by her dominant hand's palm to include multiple objects in the volume. MFB would associate the object candidates with the fingers. The user can move her palm and fingers to adjust the selection. Finally, the user flexes the corresponding finger to select the mapped, desired object. In this work, we used the index, middle, and ring fingers, and visualized the mapping lines between candidate objects and fingers in red, green, and blue, respectively, following the suggestions in [9]. In addition, we adopted the stable mapping strategy; that is, the newly entered object would take the finger assignment and its color indication from the exited object.

4 USER STUDY

This user study aims to compare and evaluate the performance and experience of the three freehand target selection techniques with two levels of target density. We followed *VR Object Selection and Manipulation Study Checklist* [3] to report our user study.

4.1 Participants and Apparatus

We recruited 18 participants (7 females, 11 males) aged between 20 to 31 ($M = 23.28, SD = 2.65$). Based on the results collected from a pre-experiment questionnaire, all participants were right-handed. Six of them had normal or corrected-to-normal vision. Twelve were familiar or very familiar with VR HMDs and used them at least once a week. Five reported being familiar or very familiar with mid-air gesture-based input.

A Meta Quest 2 was used to provide the experimental virtual environment. It has a 1832×1920 per-eye resolution, an 89° horizontal field of view, and a 120Hz refresh rate. The HMD's inside-out cameras allow 6 DoF hand tracking and finger configuration. The HMD was connected to a Win10-based desktop with an Intel i7-8700K CPU @ 3.70GHz, an NVIDIA GeForce GTX 1080 Ti GPU, and 16GB of RAM. The VR environment was created in Unity (version 2021.3.36f1c1) with Oculus Integration SDK (version 42.0).

4.2 Design, Task, and Measurements

We used a 3×2 within-subjects design with Technique (HDC vs. HCG vs. MFB) and Density (high vs. low) as the two independent variables. We adopted the target acquisition task from [26]. Several spheres (radius = 10cm) were distributed in a 1m×1m×1m cuboid space 0.8m in front of the participants. These spheres included one orange target sphere, and the remaining were white distractors. We used Poisson disk sampling [5] to randomize the positions of spherical objects, with minimum distances of 20cm and 30cm between each sphere, representing high and low density conditions, respectively (see Figure 1d-e). Participants were asked to select the orange target as fast and as accurately as possible. We applied a Latin-square design to counterbalance the order of Technique conditions and randomized the order of Density in each Technique condition. In total, we collected 2160 trials of data (= 18 participants × 3 techniques × 2 levels of density × 20 repetitions).

We collected both participants' performance data and their subjective feedback. In each trial, we recorded the *selection time*, which counted from when the objects were shown until the participants selected the target correctly. We also recorded the number of trials with correct selections—that is, successful trials. *Success rates* were then derived by dividing the number of successful trials by the total number of trials for each condition. After participants completed all trials in a Technique condition, we gave them a short version of the User Experience Questionnaire (UEQ-S) [28], a NASA-TLX questionnaire [12], and a Borg CR10 rating [4] to measure the user experience, workload, and arm fatigue of each technique. At the end of the experiment, we also asked participants to rank the techniques based on their overall preference and conducted a semi-structured interview to ask for more feedback and comments.

4.3 Procedure

The whole experiment lasted approximately 35 minutes per participant. Participants first completed a questionnaire asking about their demographic information and previous experiences with VR HMD. They were then introduced to the VR device, experimental design, and tasks. Next, they wore the HMD and started the experiment. Participants were standing while performing the tasks. Before the formal trials in each Technique condition, there was a fixed 3-minute training for participants to get familiar with the technique. After each Technique condition, they were asked to fill in the above-mentioned questionnaires, and have a rest. At the end of the experiment, they ranked the techniques and received a short interview about their subjective feelings.

5 RESULTS

5.1 Objective Measurements

We first identified and removed outliers where selection time exceeded $M \pm 3 \cdot SD$ in each condition (41 trials, ≈1.90%). Shapiro-Wilk tests and Q-Q plots indicated that both performance measures were non-normally distributed. Thus, we pre-processed the data through Aligned Rank Transform [10, 34] before performing two-way repeated-measure (RM-) ANOVA tests. Pairwise comparisons were conducted with Bonferroni corrections.

5.1.1 Selection Time. RM-ANOVA tests revealed that Technique ($F_{2,1977} = 70.587, p < .001, \eta_p^2 = .067$) and Density ($F_{1,1977} = 50.104, p < .001, \eta_p^2 = .025$) had a significant main effect on selection time. An interaction effect between Technique and Density was also found ($F_{2,1977} = 8.887, p < .001, \eta_p^2 = .010$). Figure 2a shows the significant differences in post-hoc tests. HCG ($M = 4.18s, SD = 1.51s$) was significantly slower than HDC ($M = 4.05s, SD = 2.75s$) and MFB ($M = 3.76s, SD = 1.96s$) in high density ($p < .001$ for both). Similarly, HCG ($M = 3.80s, SD = 1.44s$) was significantly slower than HDC ($M = 3.03s, SD = 1.94s; p < .001$) and MFB ($M = 3.68s, SD = 2.09s; p = .001$) in low density. Also in low density, HDC was faster than MFB ($p < .001$). The selection time by using HDC and HCG significantly increased when the density increased ($p < .001$ and $p = .007$, respectively).

5.1.2 Success Rate. Results from RM-ANOVA tests showed that Technique ($F_{2,85} = 45.3170, p < .001, \eta_p^2 = .516$) and Density ($F_{1,85} = 21.649, p < .001, \eta_p^2 = .203$) had a significant main effect on success rate. An interaction effect between Technique and Density on success rate was also found ($F_{2,85} = 5.734, p = .005, \eta_p^2 = .119$). Figure 2b summarizes results and significant differences in post-hoc tests. In high density, HCG ($M = 98.83\%, SD = 2.96\%$) had a significantly higher success rate than HDC ($M = 93.17\%, SD = 5.59\%; p < .001$) and MFB ($M = 87.44\%, SD = 8.77\%; p < .001$). In low density, both HCG ($M = 99.72\%, SD = 1.18\%$) and HDC ($M = 97.68\%, SD = 5.46\%$) had significantly higher success rates than MFB ($M = 89.33\%, SD = 7.75\%$) ($p < .001$ for both). Besides, the success rate by using HDC decreased 4.51% when the density increased ($p = .002$).

5.2 Subjective Measurements

We performed Friedman tests for subjective measurements with Technique as the only independent variable. Pairwise comparisons were also conducted with Bonferroni corrections. Figure 2c-f summarizes these results.

5.2.1 User Experience. Friedman tests showed that Technique had a significant main effect on pragmatic quality ($\chi_2^2 = 8.943, p = .011, W = .248$), but not on hedonic quality and overall quality (p=.097 and .662, respectively). Pairwise tests did not yield any significant differences on pragmatic quality ($p > .05$).

5.2.2 NASA-TLX workload. Friedman tests indicated that Technique had significant main effect on mental demands ($\chi_2^2 = 13.759$, $p = .001$, $W = .382$), physical demands ($\chi_2^2 = 11.446, p = .003, W = .318$), temporal demands ($\chi_2^2 = 12.043, p = .002, W = .335$), effort ($\chi_2^2 = 7.483, p = .024, W = .208$), frustration ($\chi_2^2 = 8.400, p =$

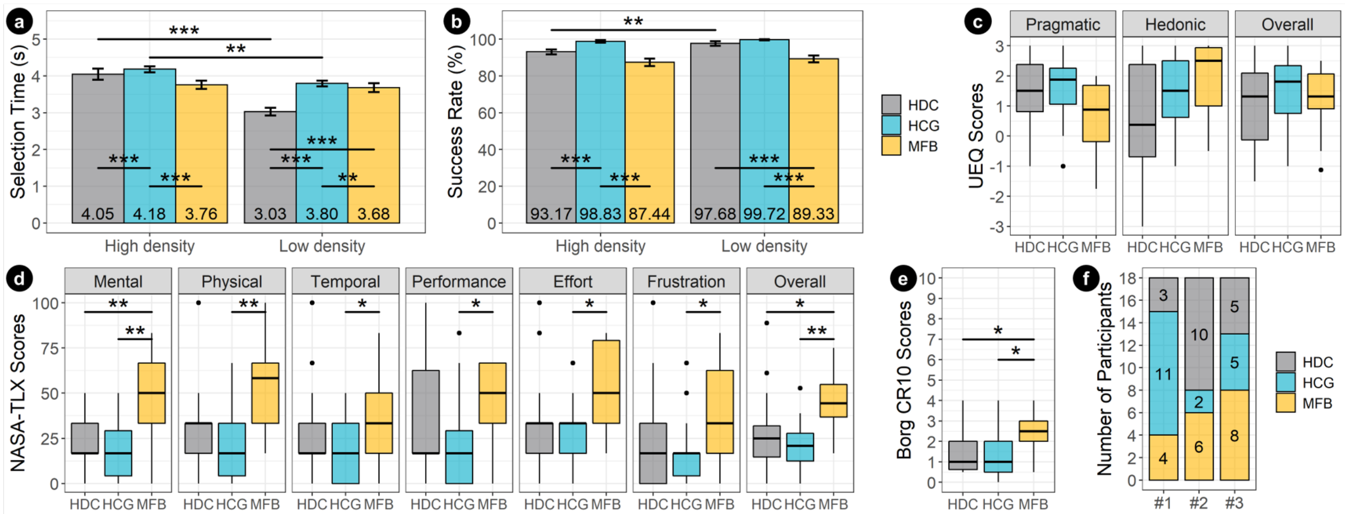

Figure 2: Plots of results: (a) bar chart of mean selection time in seconds, (b) bar chart of mean success rates in percentage, (c) boxplot of UEQ scores (higher scores are better), (d) boxplot of NASA-TLX scores (lower scores are better), (e) boxplot of Borg CR10 scores (lower scores are better), (f) users' rankings (#1, #2, and #3) of the three techniques. The error bars in (a-b) represent standard errors. Significant differences in pairwise comparisons are marked with *, **, and *, representing a Bonferroni-adjusted significance level of .05, .01, and .001, respectively.**

.015, $W = .233$), and overall workload ($\chi^2_2 = 14.111, p < .001, W = .392$). A close significant effect of Technique on Performance ($\chi^2_2 = 5.719, p = .057, W = .159$) was found. Significant differences identified from pairwise tests were summarized in Figure 2d. Particularly, MFB ($Mdn = 44.44$) led to a significantly higher workload than HDC ($Mdn = 25.00; p = .021$) and HCG ($Mdn = 20.83; p = .003$).

5.2.3 Arm Fatigue. For perceived arm fatigue collected via Borg CR10 questionnaire, there was a significant main effect of Technique ($\chi^2_2 = 10.714, p = .005, W = .298$). Results from pairwise comparisons showed that MFB ($Mdn = 2.50$) induced more arm fatigue than HDC ($Mdn = 1.00; p = .032$) and HCG ($Mdn = 1.00; p = .026$).

5.2.4 Overall Preference. As shown in Figure 2f, participants' preferences varied. 11 participants (61%) ranked HCG as the most favored technique, and 10 (55.56%) ranked HCG as the second. For the third place (the least favored technique among the three), 5 participants (27.28%) chose HDC, 5 chose (27.28%) HCG, and 8 (44.44%) chose MFB.

6 DISCUSSION

6.1 Technique Evaluation

6.1.1 HandDepthCursor (HDC). More than half ($N = 10$) of participants ranked HDC in second place. In the interview, participants mentioned that the forward/backward move gestures were easy to learn and use, and collaborated well with the raycasting technique. There were two main weaknesses of HDC mentioned by participants. First, four participants complained the moving speed of the depth cursor was too slow. When an object was partially occluded at a distance, they preferred to select the target directly rather than disambiguate the cluttered environment. This led to a dropped success rate in high density conditions where objects were closer to each other (see Figure 2b). However, other participants felt the speed was suitable. Thus, we recommended offering users the freedom to customize the cursor's moving speed if HDC is used in real applications, like the settings on desktops. Second, similar to other depth-based techniques, HDC reduces the disambiguation in the depth axis but not the x-y plane. When such a case happens at a distance, participants felt that is challenging to maintain the pointing gesture, resulting in a jitter among the objects. On the whole, HDC is fast, accurate, and highly usable (see Figure 2), especially in a relatively low density environment.

6.1.2 HandConeGrid (HCG). Overall, HCG outperformed the other two techniques in our study. Though it was statistically slower than the other two (see Figure 2a), participants did not feel it was slow in practical use. Participants did not give explicit comments on the hand gestures. While P1 and P8 mentioned that they were "*highly involved in the selection process without extra physical efforts*". We believe HCG, including its mechanism and the proposed gestures, involves a smaller motor space, thereby minimizing their effort to perform precise target selection in dense environments. However, participants suggested not including the area selection phase when they could select the target (e.g., the target was positioned in front of them). HCG involves area selection and target selection, both of which use the same confirmation mechanism (i.e., a pinch gesture using the dominant hand). One solution can be using different gestures for the two types of selection. Another solution is integrating a suitable mode-switching action for activating the area selection [29, 31].

6.1.3 MultiFingerBubble (MFB). MFB did not perform well on the whole. Its success rate was the lowest among the three in either level of density. It is worth mentioning that MFB was the only technique having a similar selection time and success rate in two levels of density, which means the density may not affect its selection process much. On the other hand, MFB led to a significantly higher workload and more arm fatigue than the other two. Seven participants reported unintended activations (due to unintentional finger flexes), which were not reported or described in detail by [9]. One possible reason is that we used inside-out cameras on the headset to track hand gestures rather than tracking gloves. When participants reached out their hands for pointing to the target, the cameras may not configure their fingers accurately because they might be blocked by the wrist or forearm. MFB was a 'controversial' technique according to participants' preferences (see Figure 2f) and feedback. The four participants who ranked MFB as the most preferred technique felt it was "novel" and "interesting". However, the others highlighted its high learning costs. Besides, six participants mentioned they were not able to flex their middle or ring fingers while holding the others stable smoothly, which eventually led to extra caution.

6.2 Lessons Learned

Based on the study results, we extracted three lessons (**#L**) for designing bare-hand mid-air pointing techniques.

- **L1**. HandDepthCursor and HandConeGrid are suggested due to their higher accuracy, lower workload, and arm fatigue. When the environmental density is relatively high, HandConeGrid can be the first choice, given its powerful disambiguation mechanism and efficient gestural operations.
- **L2**. The designed bare-hand mid-air pointing technique may inevitably need the same action for multiple purposes, such as a pinch gesture for all types of selection confirmation. In such a case, a suitable mode-switching mechanism should be designed and integrated into the technique.
- **L3**. Designers should consider the capabilities of hand-tracking devices. In this study, we found finger actions might be blocked by other parts of the hand, which affected user performance.

6.3 Limitations and Future Work

As a first exploration of freehand pointing techniques for dense environments in VR HMDs, this work has two limitations, which represent possible avenues for future work. First, the MultiFingerBubble [9] was first proposed and implemented using haptic gloves, while we used the headset's built-in cameras to track the hands as a bare-hand approach. The difference in tracking approach may have led to slightly different results. Second, based on the taxonomy [1, 42], we proposed two techniques with defined gestures. In the future, we want to explore other possible gestures for the proposed techniques, enhance the techniques, and compare them with the controller-based techniques. For example, recently, Yu et al. [41] proposed design patterns that combined on-boy and mid-air interfaces for VR interaction. Their work opens more opportunities for designing mid-air selection techniques in dense VR environments. We plan to explore the possibility of combining bare-hand and on-body techniques with suitable disambiguation mechanisms in the future. In addition, we plan to test further optimized techniques in real VR scenarios and with other properties of the objects (e.g., non-regular, arbitrary shapes) and mixed with non-selectable environmental objects as distractors.

7 CONCLUSION

In this paper, we explored bare-hand mid-air pointing selection techniques for dense VR environments. We first reviewed existing controller-based techniques and developed two gestural adaptions—HandDepthCursor and HandConeGrid. We then compared and evaluated them with an existing technique, MultiFingerBubble, via a user study using two levels of density. Our results showed that HandDepthCursor and HandConeGrid were fast and accurate. HandConeGrid had particular advantages in densely cluttered environments. Based on our results, we summarized three lessons for future design and development of bare-hand mid-air techniques for target selection in dense VR environments.

ACKNOWLEDGMENTS

We thank the volunteers who participated in the user study. We also thank the reviewers for their valuable time and insightful comments that helped improve our paper. This research was partly funded by Xi'an Jiaotong-Liverpool University Special Key Fund (#KSF-A-03), the National Science Foundation of China (#62272396), and Suzhou Municipal Key Laboratory for Intelligent Virtual Engineering (#SZS2022004).

REFERENCES

[1] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (2013), 121–136. https://doi.org/10.1016/j.cag.2012.12.003

[2] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: A 3D Pointing Facilitation Technique Based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300331

[3] Joanna Bergström, Tor-Salve Dalsgaard, Jason Alexander, and Kasper Hornbæk. 2021. How to Evaluate Object Selection and Manipulation in VR? Guidelines from 20 Years of Studies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 533, 20 pages. https://doi.org/10.1145/3411764.3445193

[4] Gunnar Borg. 1982. Psychophysical bases of perceived exertion. *Medicine and science in sports and exercise* 14, 5 (1982), 377–381.

[5] Robert Bridson. 2007. Fast Poisson Disk Sampling in Arbitrary Dimensions. In *ACM SIGGRAPH 2007 Sketches* (San Diego, California) *(SIGGRAPH '07)*. Association for Computing Machinery, New York, NY, USA, 22–es. https://doi.org/10.1145/1278780.1278807

[6] Gavin Buckingham. 2021. Hand Tracking for Immersive Virtual Reality: Opportunities and Challenges. *Frontiers in Virtual Reality* 2 (2021), 6 pages. https://doi.org/10.3389/frvir.2021.728461

[7] Jeffrey Cashion, Chadwick Wingrave, and Joseph J. LaViola Jr. 2012. Dense and Dynamic 3D Selection for Game-Based Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 18, 4 (2012), 634–642. https://doi.org/10.1109/TVCG.2012.40

[8] Gerwin de Haan, Michal Koutek, and Frits H. Post. 2005. IntenSelect: Using Dynamic Object Rating for Assisting 3D Object Selection. In *Proceedings of the 11th Eurographics Conference on Virtual Environments* (Aalborg, Denmark) *(EGVE'05)*. Eurographics Association, Goslar, DEU, 201–209.

[9] William Delamare, Maxime Daniel, and Khalad Hasan. 2022. MultiFingerBubble: A 3D Bubble Cursor Variation for Dense Environments. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 453, 6 pages. https://doi.org/10.1145/3491101.3519692

[10] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 754–768. https://doi.org/10.1145/3472749.3474784

[11] Tovi Grossman and Ravin Balakrishnan. 2006. The Design and Evaluation of Selection Techniques for 3D Volumetric Displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology* (Montreux, Switzerland) *(UIST '06)*. Association for Computing Machinery, New York, NY, USA, 3–12. https://doi.org/10.1145/1166253.1166257

[12] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909

[13] Regis Kopper, Felipe Bacim, and Doug A. Bowman. 2011. Rapid and accurate 3D selection by progressive refinement. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, Singapore, 67–74. https://doi.org/10.1109/3DUI.2011.5759219

[14] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional, Boston, MA, USA.

[15] Yang LI, Jin HUANG, Feng TIAN, Hong-An WANG, and Guo-Zhong DAI. 2019. Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware* 1, 1 (2019), 84–112. https://doi.org/10.3724/SP.J.2096-5796.2018.0006

[16] Jiandong Liang and Mark Green. 1994. JDCAD: A highly interactive 3D modeling system. *Computers & graphics* 18, 4 (1994), 499–506.

[17] Xueshi Lu, Difeng Yu, Hai-Ning Liang, and Jorge Goncalves. 2021. iText: Hands-Free Text Entry on an Imaginary Keyboard for Augmented Reality Systems *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 815–825. https://doi.org/10.1145/3472749.3474788

[18] Yiqin Lu, Chun Yu, and Yuanchun Shi. 2020. Investigating Bubble Mechanism for Ray-Casting to Improve 3D Target Acquisition in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Atlanta, GA, USA, 35–43. https://doi.org/10.1109/VR46266.2020.00021

[19] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. 2019. A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments. *Computer Graphics Forum* 38, 1 (2019), 21–45. https://doi.org/10.1111/cgf.13390 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13390

[20] Xuanru Meng, Wenge Xu, and Hai-Ning Liang. 2022. An Exploration of Hands-free Text Selection for Virtual Reality Head-Mounted Displays. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 74–81. https://doi.org/10.1109/ISMAR55827.2022.00021

[21] Mark R Mine. 1995. Virtual environment interaction techniques. *UNC Chapel Hill CS Dept* (1995), 18 pages.

[22] Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2021. Pinch, Click, or Dwell: Comparing Different Selection Techniques for Eye-Gaze-Based Pointing in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 15, 7 pages. https://doi.org/10.1145/3448018.3457998

[23] A. Olwal, H. Benko, and S. Feiner. 2003. SenseShapes: using statistical geometry for object selection in a multimodal augmented reality. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings*. IEEE, Tokyo, Japan, 300–301. https://doi.org/10.1109/ISMAR.2003.1240730

[24] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[25] Yuan Yuan Qian and Robert J. Teather. 2017. The Eyes Don't Have It: An Empirical Comparison of Head-Based and Eye-Based Selection in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 91–98. https://doi.org/10.1145/3131277.3132182

[26] Jonas Schjerlund, Kasper Hornbæk, and Joanna Bergström. 2021. Ninja Hands: Using Many Hands to Improve Target Selection in VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 130, 14 pages. https://doi.org/10.1145/3411764.3445759

[27] G. Schmidt, Y. Baillot, D.G. Brown, E.B. Tomlin, and J.E. Swan. 2006. Toward Disambiguating Multiple Selections for Frustum-Based Pointing. In *3D User Interfaces (3DUI'06)*. IEEE, Alexandria, VA, USA, 87–94. https://doi.org/10.1109/VR.2006.133

[28] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 6 (2017), 103–108. https://doi.org/10.9781/ijimai.2017.09.001

[29] Rongkai Shi, Nan Zhu, Hai-Ning Liang, and Shengdong Zhao. 2021. Exploring Head-based Mode-Switching in Virtual Reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Bari, Italy, 118–127. https://doi.org/10.1109/ISMAR52148.2021.00026

[30] Frank Steinicke, Timo Ropinski, and Klaus Hinrichs. 2006. *OBJECT SELECTION IN VIRTUAL ENVIRONMENTS USING AN IMPROVED VIRTUAL POINTER METAPHOR*. Springer Netherlands, Dordrecht, 320–326. https://doi.org/10.1007/1-4020-4179-9_46

[31] Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2019. Experimental Analysis of Barehand Mid-Air Mode-Switching Techniques in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300426

[32] Lode Vanacken, Tovi Grossman, and Karin Coninx. 2007. Exploring the Effects of Environment Density and Target Visibility on Object Selection in 3D Virtual Environments. In *2007 IEEE Symposium on 3D User Interfaces*. IEEE, Charlotte, NC, USA, 8 pages. https://doi.org/10.1109/3DUI.2007.340783

[33] Yushi Wei, Rongkai Shi, Difeng Yu, Yihong Wang, Yue Li, Lingyun Yu, and Hai-Ning Liang. 2023. Predicting Gaze-based Target Selection in Augmented Reality Headsets based on Eye and Head Endpoint Distributions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, 14 pages. https://doi.org/10.1145/3544548.3581042

[34] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[35] Wenge Xu, Hai-Ning Liang, Anqi He, and Zifan Wang. 2019. Pointing and Selection Methods for Text Entry in Augmented Reality Head Mounted Displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 279–288. https://doi.org/10.1109/ISMAR.2019.00026

[36] Wenge Xu, Hai-Ning Liang, Yuxuan Zhao, Tianyu Zhang, Difeng Yu, and Diego Monteiro. 2019. RingText: Dwell-free and hands-free Text Entry for Mobile Head-Mounted Displays using Head Motions. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (2019), 1991–2001. https://doi.org/10.1109/TVCG.2019.2898736

[37] Wenge Xu, Xuanru Meng, Kangyou Yu, Sayan Sarcar, and Hai-Ning Liang. 2022. Evaluation of Text Selection Techniques in Virtual Reality Head-Mounted Displays. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 131–140. https://doi.org/10.1109/ISMAR55827.2022.00027

[38] Difeng Yu, Hai-Ning Liang, Kaixuan Fan, Heng Zhang, Charles Fleming, and Konstantinos Papangelis. 2020. Design and Evaluation of Visualization Techniques of Off-Screen and Occluded Targets in Virtual Reality Environments. *IEEE Transactions on Visualization and Computer Graphics* 26, 9 (2020), 2762–2774. https://doi.org/10.1109/TVCG.2019.2905580

[39] Difeng Yu, Hai-Ning Liang, Feiyu Lu, Vijayakumar Nanjappan, Konstantinos Papangelis, and Wei Wang. 2018. Target Selection in Head-Mounted Display Virtual Reality Environments. *J. Univers. Comput. Sci.* 24 (2018), 1217–1243.

[40] Difeng Yu, Hai-Ning Liang, Xueshi Lu, Kaixuan Fan, and Barrett Ens. 2019. Modeling Endpoint Distribution of Pointing Selection Tasks in Virtual Reality Environments. *ACM Trans. Graph.* 38, 6, Article 218 (nov 2019), 13 pages. https://doi.org/10.1145/3355089.3356544

[41] Difeng Yu, Qiushi Zhou, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2022. Blending On-Body and Mid-Air Interaction in Virtual Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Singapore, Singapore, 637–646. https://doi.org/10.1109/ISMAR55827.2022.00081

[42] Difeng Yu, Qiushi Zhou, Joshua Newn, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. Fully-Occluded Target Selection in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3402–3413. https://doi.org/10.1109/TVCG.2020.3023606