Ninja Hands: Using Many Hands to Improve Target Selection in VR

Jonas Schjerlund
josh@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

Kasper Hornbæk
kash@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

Joanna Bergström
joanna@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

Figure 1: A sample VR scene with *Ninja Hands*. It maps one physical hand to many distributed virtual hands, allowing the user to comfortably reach distant objects.

ABSTRACT

Selection and manipulation in virtual reality often happen using an avatar's hands. However, objects outside the immediate reach require effort to select. We develop a target selection technique called *Ninja Hands*. It maps the movement of a single real hand to many virtual hands, decreasing the distance to targets. We evaluate Ninja Hands in two studies. The first study shows that compared to a single hand, 4 and 8 hands are significantly faster for selecting targets. The second study complements this finding by using a larger target layout with many distractors. We find no decrease in selection time across 8, 27, and 64 hands, but an increase in the time spent deciding which hand to use. Thereby, net movement time still decreases significantly. In both studies, the physical motion exerted also decreases significantly with more hands. We discuss how these findings can inform future implementations of the Ninja Hands technique.

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445759

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Interaction techniques**.

KEYWORDS

virtual reality, many hands, user studies, target selection, virtual hands

ACM Reference Format:
Jonas Schjerlund, Kasper Hornbæk, and Joanna Bergström. 2021. Ninja Hands: Using Many Hands to Improve Target Selection in VR. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3411764.3445759

1 INTRODUCTION

In virtual reality (VR), users typically use the hands of their avatar to select and manipulate virtual objects. This provides an intuitive way to interact with the virtual environment. Consequently, much research in VR has been about how to represent avatar hands [27, 31, 34, 39], how to control avatar hands [37, 48], and how to provide haptic feedback when avatar hands touch virtual objects [3].

However, selecting objects outside of arm's reach is slow. Actions are added to the selection task when users have to walk to the target, teleport their avatar near the target [9], or bring the target close to the hand reach by room-scaling [35]. Some interaction techniques use ray-casting [37, 50] or change the movement gain of the hand [36] to enable reaching into far space. While useful, this makes controlling distant movement inaccurate or effortful.

An alternative to mimicking the user's real hands in VR is to map them to many virtual hands. For desktop computers, *ninja cursors* [28] allowed the user to control many cursors simultaneously. This type of control significantly lowered target selection times compared to controlling only a single cursor. We hypothesize that a similar approach could work in VR. It could allow reaching far while minimizing physical motion. Further, it could support hand-based selection and interaction across large virtual worlds. However, it is unclear if avatar hands can retain their intuitiveness and effectiveness if multiplied and distributed.

To address these questions, we present the *Ninja Hands* technique for improved target selection in VR. *Ninja Hands* maps the movement of many virtual hands to that of a single physical hand (Figure 1). We report on two empirical studies that explore the performance of the technique and the users' experience. The first study suggests that using 4 and 8 *Ninja Hands* in a 2.5m×2.5m×2.5m section of space lowers target selection times, reduces physical effort by minimizing movement, and gives comparable subjective satisfaction to using a single hand. The second study shows that given the same task in a larger 10m×5m×10m environment, 8, 27, and 64 hands perform comparably in target selection time. However, how this time is spent changes with the hand count; at higher hand counts, the time spent deciding which hand to use increases, which leads to a corresponding decrease in time spent moving. Participants still move significantly less with higher hand counts. We also find that higher hand counts result in a lower overall workload, but this benefit appears to reverse at higher hand counts, suggesting the limitations of increasing the number of hands. Together, these results illustrate the benefits and drawbacks of a new way of interacting in VR and suggest that avatars that differ from our physical bodies are useful, controllable, and enjoyable.

2 RELATED WORK

Ninja Hands uses many hands for target selection in VR. Here, we discuss the use of avatar hands in VR, how the hands can reach distant targets, and how many hands could aid performance in target selection.

2.1 Virtual Hands

Virtual hands have been extensively used in VR research since the 1980s [22, 37]. Recent consumer-level VR devices, such as the Valve Index and Oculus Quest, also offer integrated support for hand tracking, as well as virtual controller models that show a hand holding the controller and touching the same buttons as the user.

The use of avatar hands offers many benefits in inducing positive user experiences and improving task performance. Many of these benefits depend on the visual appearance and tracking accuracy of the hands. For example, the visual appearance and realism of avatars have been shown to affect presence, the subjective experience of "being there" [44]; body ownership, the subjective experience of ownership over the virtual avatar's body [43]; and the overall embodiment of virtual hands [5]. Similarly, presence and body ownership results specifically for virtual hands indicate that more realistic hands outperform abstract or non-human hand representations such as flat-textured hands, spheres, tracking points, or none at all [2, 27, 39, 40]. Further, it has been shown that this sense of ownership can be retained even if the virtual hand is displaced from the physical one [13]. Users of more realistic hand representations also outperform less realistic or unrepresented hands in performance-based tasks such as typing [27] or pointing [41]. However, the drawback of realistic representations of hands is that they are similarly limited by the arm's reach of the user.

2.2 Reaching Out in VR

One technique to increase reach beyond physical motions is to use movement gains. Examples include manipulating the control-display ratio, for example between a mouse and a cursor or a VR controller and the virtual hand. In VR, for example, a classic technique to do this with avatar hands is the *Go-go* technique [36]; when the user reaches beyond a certain threshold, instead of following their physical hand, the virtual hand extends exponentially further.

Another technique is to increase the selection area into distant parts of the virtual environment. The most common approach for this is ray-casting, both for mid-air interaction, large screens, and VR. In VR, ray-casting is typically based on the pointing direction of a hand-held controller (e.g., [8, 37, 50]), though the head has also been as the origin [1, 47]. It has further been combined with hands by spawning them at the target selected by the ray-cast [8]. However, ray-casting faces issues in selecting occluded objects and precision at a distance [20, 37].

A similar approach is to extend the virtual arm into the distance [12, 26]. Here the arm does not extend to infinity as a ray does, but it is unclear how to control the length of the arm effectively. Thereby, it is also unclear how effectively targets that are very far from or close to the user can be reached.

Some techniques also bring the distant parts of the environment within the user's reach. A classic example of this is the World-in-Miniature [45], wherein a miniature version of the larger virtual space can be used to reach the virtual objects at a distance or occluded objects. Similarly, [11] present a technique that brings distant parts of the room within the real reach, but by down-scaling the depth of the virtual room.

Finally, one class of techniques increases the selection area (e.g., [15, 48]). This has been demonstrated, for instance, with a 3D *bubble cursor* [48], volumetric cursor [51], and with cones [15]. The challenge with these techniques is increasing precision in selecting single targets when their area covers multiple targets, as well as at great distances. Recent work that compares multiple several of these techniques [30] shows that by augmenting them with a targeting mechanism inspired by the *bubble cursor* [48], these issues can be alleviated. However, doing so involves abstract pointing metaphors that is far removed from natural hand motion and behavior.

2.3 Mapping from One to Many

Instead of increasing the selection area or distance of a single hand, previous work on VR has also suggested using multiple limbs. For example, Hoyet et al. [24] gave their participants an additional sixth finger, and Guterstam et al. [17] a third arm. However, these previous works focus on the experience of having them, such as body acceptance and ownership. It is unclear whether they could help improving performance in interactive tasks.

Ninja cursors [28] suggests they could. It maps input from a single mouse to *n* virtual cursors distributed across a desktop display to improve target acquisition efficiency for large 2D displays. A single mouse synchronously moves multiple cursors on the monitor. Only one cursor can actively hover over a target at a time; if a cursor hovers over a target while another is active, it is stopped in place and added to a queue of waiting cursors until the active cursor stops hovering over the target, in which case the next cursor in the queue is made active based on a first-in-first-out principle. Later work has expanded the technique by introducing an additional input modality, gaze tracking. This is seen in *rake cursor* [6], where gaze tracking is used to choose which cursor in a grid that is active. The work of Räihä and Špakov [38] similarly uses gaze tracking for disambiguation. In VR, Lubos et al. [31] use head tracking to disambiguate between two sets of virtual hands. However, they did not investigate whether manipulating the number of hands can improve shortest-distance gains, but instead focused on control-sharing between the hands. A key finding in this body of work is that having many effectors (i.e., cursors or hands) improves efficiency, with an additional workload associated with many effectors [6, 28, 38].

Thereby, we summarise that the use of virtual hands is beneficial for user experience in VR. The positive user experience seems to transfer for changed bodies, such as to extended arms [12] or to multiple limbs [17]. Based on the findings about the *ninja cursors* technique, mapping from one hand to many hands could improve target selection performance by alleviating some of the drawbacks that other reaching techniques pose. However, this may decrease the user experience. To investigate if and when mapping the real hand to many virtual hands could improve target selection performance, we design a target selection technique: *Ninja Hands.*

3 THE TECHNIQUE

The *Ninja Hands* target selection technique enables many hands in VR. Three interrelated parameters determine how one hand can control multiple hands at once. One, the number and arrangement of the hands; two, the mapping that defines how they move relative to the physical hand; and three, the disambiguation that determines which hand should be active if multiple hands touch targets simultaneously. This section outlines the design space of these parameters.

3.1 Number and arrangement

The *number* of virtual hands in *Ninja Hands* is two or more, and affects their *arrangement.* Given an even distribution of targets and *n* hands in a virtual environment, this will reduce the shortest distance from a target to a hand by a factor of $\sqrt{n}$ (similarly to the theoretical benefit of *ninja cursors* [28]).

The number of hands is tied to their arrangement and determines its functionality. For example, a line of two hands spaced across a room allows simultaneous selection of objects close to and far from the user. Increasing the number of hands in the line allows for easy access to objects that are spaced across a surface (e.g., a counter) that extends away from the user. Similarly, an arrangement of eight (2^3) hands can form the corners of a cube well suited for, for example, interacting with objects spaced apart in the corners of a room. Maintaining this cubic arrangement but increasing the number to, for example, 27 (3^3) hands create a distribution that could be well suited for a room with many clusters of objects that are spaced apart. The arrangement can also vary in scale; for example, a smaller cube of 27 hands can form a volumetric selection volume (similar to the *Silk Cursor* [52]). Arrangements might also take on more complex shapes for more specialized tasks; for example, we envision arm-shaped arrangements of hands that extend into the room, allowing the user to interact with anything that touches their extended arm, or smaller clusters of hands that align with irregularly distributed clusters of objects.

The main trade-off for the number and arrangement of hands is occlusion and decision-making efforts versus shortest-distance gains. Increasing the number of hands can cause a cluttering effect where hands occlude objects and other hands, or objects occlude hands. Similarly, previous work [6, 38] has suggested that having to choose from many cursors will lead to an increase in cognitive load at higher numbers, and we speculate this also holds for many hands. However, by increasing the number of hands, we also reduce the shortest distance to any given target, making targets more comfortable to reach. Another trade-off is specialization versus generalizability. It is possible to specialize the arrangement and number of hands for a given environment, so that for example, each interactable object in a scene has an adjacent hand. However, this means that the given number and arrangement cannot be generalized to any environment, which, for example, an arrangement of evenly distributed hands might.

In our studies, we investigate how hand count affects the effectiveness and usability of generic arrangements, similar to work that has been done in 2D desktop environments (e.g., [6, 28, 38]). Further, we wish to investigate the trade-offs associated with increasing the number of hands. The first study features four hands arranged in a grid and eight hands in a cube, whereas the second study extends the cubic arrangement to the scale of a large room and features 8, 27, and 64 hands.

3.2 Mapping

Mapping describes the movement-display ratio for the virtual hands. It defines how the virtual hands move when the user moves their physical hand.

The mapping can be constant or based on an algorithm. A constant mapping means that the virtual hands move a set distance based on the physical hand's movements, as in the *Silk Cursor* [51], where the movements of the physical hand maps directly to the volumetric cursor; potentially, a scaling factor can be applied to increase reach. An algorithm-based mapping means that some factor in the mapping changes the relationship between the physical and virtual hands, such as in the *Go-go* technique [36], where the virtual hand moves exponentially faster when extended beyond a predetermined threshold.

The main trade-off for the mapping is coverage against precision. This is largely informed by the number and arrangement of the hands. For example, if a low number of hands are arranged in the center of the room, a constant mapping will not allow them to reach the corners of the room, unless a scaling factor is applied. An exponential mapping will, but at the cost of a loss of precision when reaching the edges of the room. However, one of the main advantages of the *Ninja Hands* technique is that by manipulating the arrangement and number of hands, we can decrease the need for extreme mapping functions that cause this loss of precision. For example, evenly distributing a larger number of hands in the room means that anywhere can be comfortably reached with a small scaling factor or minimal linear function to increase the reach of the hands. Similarly, the user's reach could also be computed and the room filled with exactly enough hands to allow anywhere to be reached with a constant mapping and no scaling factor.

We do not wish to disproportionately advantage the more-handed conditions by applying a variable mapping based on the hand count. Therefore, we choose a constant gain with a scaling factor in our studies. This scaling factor is determined so that the single hand can reach all targets and is then applied to all hands. The particular scaling factor for each study is reported in its design subsection.

3.3 Disambiguation

In the *Ninja Hands* technique, several hands may touch targets at once; disambiguation concerns how to decide which virtual hand should be active. We make a fundamental assumption that there should only be one active hand at a time; this allows the user to better focus on interacting with a single object.

There are two established approaches to disambiguation. One is a queue-based algorithm, as in *ninja cursors* [28]. In *ninja cursors*, only one cursor can actively touch a target. If other cursors touch targets while one is active, they are frozen in place and added to a queue, and when the active cursor stops touching its target, cursors in the queue are made active based on a first-in-first-out principle. In the edge case of multiple cursors reaching a target simultaneously, the cursor closest to the center of the target is made active. The second option is gaze tracking, where the convergence of the user's gaze determines which hand should be active (as in [6, 38]).

The importance of disambiguation is determined by the ratio between the number of hands, their arrangement, and the number and arrangement of objects in the scene. The main advantage of the queue-based approach is the generalizability; it can function in any virtual environment and with any VR hardware configuration. Meanwhile, though gaze tracking can more quickly cycle through many hands, it will struggle in busy scenes where gaze might be occluded, and gaze-tracking hardware currently has limited availability.

We seek to understand how *Ninja Hands* functions in its most generalizable form. Therefore, we implement the queue-based approach for disambiguation in our studies. Since all targets and hands in our studies are equally big, in the edge case that two hands reach a target simultaneously, a random one is chosen to be active.

4 FIRST STUDY

This study aims to evaluate the target selection performance and experience of using the *Ninja Hands* technique. The study uses a target acquisition task featuring two implementations of the technique, as well as a single hand with identical gain as a control condition. These hand arrangements are tested with a high and low density of distractors to establish if performance is equivalent given different densities of targets within the same, relatively small, space.

4.1 Design

The study was within-subjects and had two independent variables: *Hand arrangement* (one hand, four hands in a grid, eight hands in a cube) and *distractor density* (low, high).

The *hand arrangements* were a single hand compared against two numbers and arrangements of *Ninja Hands*. One hand represented a typical long-distance VR selection technique. Four hands arranged in a grid represented a two-dimensional arrangement with no depth variation. Eight hands arranged in a cube extended this concept into three dimensions, introducing depth variation. These configurations featured the same constant mapping with a scaling factor of two (roughly equivalent to the difference between the bounds of a comfortable 50cm reach and the bounds of the target space).

The hands moved and rotated based on the user's physical hand motion. To alleviate potential biases associated with the appearance of the hands, they were implemented wearing dark gloves (see Figure 7). Basic finger tracking was implemented using the capacitive sensors in the SteamVR Knuckles DV controllers, to increase the experience that they were virtual replicas of the user's hand.

The study used a target acquisition task, wherein the user must use different hand configurations to select eight targets in a 2.5m ×2.5m×2.5m target space with varying densities of distractors (see Figure 2). The targets were eight red spheres (15cm diameter) distributed in a 2.0m×2.0m×2.0m cube within this space. Spherical targets have been used in similar previous research and present an easy way to control target size in all dimensions at variable distances close to the user [31, 46]. The space was populated with visually identical distractors. The *distractor density* is defined as the minimum distance between each target. The distractor density was $\frac{1}{4}$th and $\frac{1}{3}$rd of the target space for high and low density, respectively; we further subtracted the target diameter. This made high-density 47.5cm and low-density 68.3cm. To place the targets, we used Poisson disk sampling [10] with these values to guarantee

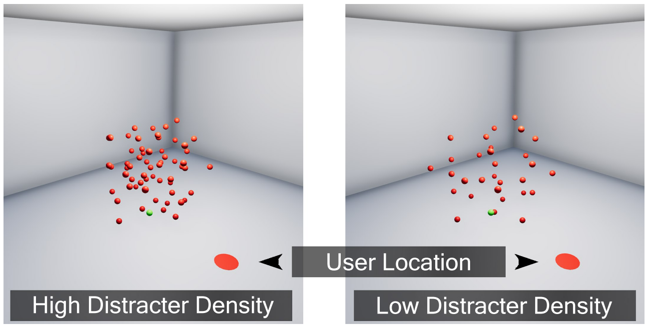

Figure 2: The 10m×5m×10m virtual environment seen from the top corner of the room. The targets are distributed within a 2.5m×2.5m×2.5m subsection of the environment. The user stands on the red mark on the floor. For scale reference, the red mark has a 50cm diameter.

the minimum distance between them. The user stood 2m from the target space, on a red spot marked on the floor. These configurations can be seen in Figure 2. This approach to target generation ensures a variety of depths and angular sizes for hands and targets.

The participant performed 8×12 repetitions 96 per hand arrangement and distractor density for 576 observations per participant (96 trials × 3 hand arrangements × 2 distractor densities). The study was balanced using a Latin square for the hand arrangements, which was repeated twice. In the first Latin square, the participants first have the high distractor density for each hand condition, then the low; the second Latin square was low first for each hand condition, then high. This gave us a total of six permutations, which was repeated three times for the 18 participants.

4.2 Measures

We report the completion time for each trial, defined as the time in seconds from when the user presses a controller button to start the trial until they press the same button to make a selection. We also report physical motion, defined as the distance in centimeters that the controller is moved during a trial. Lastly, we report error rates, an error defined as when the user selects a distractor instead of a target.

To examine whether the subjective experience of efficiency correlates with our quantitative measures, we also include a questionnaire. We use an 11-question user satisfaction questionnaire (normalized to a 7-point Likert scale) used in previous work [23], adapted from QUIS [42] and the ISO 9241-9 standard for pointing devices.

4.3 Apparatus

We used an HTC Vive VR HMD with a display resolution of 2160×1200, 90 HZ refresh rate, and a 110° field of view. Prototype SteamVR Knuckles DV controllers were used as hand-held input devices.

We powered the VR application with a Windows 10-based PC with an NVidia GeForce 1070 GPU, an Intel i7-8750H CPU @ 2.2GHz, and 32 GB of DDR4 RAM. The VR application was implemented using Unity (version 2019.1.10f) and the SteamVR Unity plugin (version 2.3.2).

4.4 Participants

We recruited 18 participants (9 female), age range 21-35, mean = 27, SD = 4.07. To control the potential influences of handedness, we only recruited right-handed participants. As we use a red-green target scheme, we also excluded persons with color blindness from the study. Participants were required to pass a color blindness test online when signing up for participation. The test was re-administered on-site before running the study.

Participants were recruited through social media groups and email lists for people interested in participating in scientific experiments. Each participant received the equivalent of €15 for their time.

4.5 Procedure

Upon completing the color-blindness test, we explained the purpose of the experiment to the participant, had them sign an informed consent form, and placed them into VR. The participant was instructed to move the virtual hands with their right hand and was allowed to move the virtual hands around and observe the controllers' finger tracking before the first trial was started.

The participant stood 2m away from the bounds of the targets on a fixed spot marked on the floor. The full arrangement of targets was within the field of view of the participant. This configuration kept the hand activity within the participant's field of view.

Before each trial, there was a reset step where the participant must resume a default waist-level resting position while the hands were not shown. Only when the participant pressed the trigger would the next trial begin. This follows previous study designs [28] and alleviates potential motion bias from hand placement after selection.

When a trial began, the virtual hands appeared centered in the target space, and the intended target became green. All other spheres were red and served as distractors for this trial. The participant was tasked with moving any hand to the intended target and selecting it by pressing the controller's trigger. Each hand and target has a spherical collider which encapsulates it. Targets are selected when a hand and target collider intersect and the user presses a trigger on the controller. The hand size is the default hand size in the SteamVR framework, and targets have a 15cm diameter. Though the size of the hand will influence selection, the impact is consequently balanced across participants.

After 96 selections per distractor density, the hand condition ended. The participant filled out a questionnaire outside of VR, had a one minute break, and started the next hand condition.

5 FIRST STUDY RESULTS

This section describes the outcome of the first study. We first discard the first round of 8 trials for each condition to balance initial training effects (8.33% of the total number of trials). We then discard outliers, defined as data points that fall outside 1.5 times the interquartile range for completion time and motion. 4.6% outliers were discarded for completion time and 2.8% outliers were discarded for motion. We take these to represent trials where some external factor, such as momentary loss of tracking, caused participants to idle or move far more than they otherwise would.

5.1 Completion Time

Completion time is defined as the time in seconds it took for a participant to complete one trial successfully. This time is calculated from the difference between the moment when the user ends the resting phase by pressing the trigger and the moment when the user successfully selects the target. Figure 3 shows the mean completion time per trial. The overall mean completion times were 1.51s for the single hand in high density and 1.48s in low density, 1.45s for the grid in high density and 1.38s in low density, as well as 1.43s for the cube in high density and 1.30s for low density. We performed two-way repeated measures ANOVAs for the trial duration, hand arrangement and distractor density. This revealed a significant effect of hand arrangement ($F_{2,34} = 3.56, p < .04$) and distractor density ($F_{1,17} = 7.59, p < .02$). There were no interaction effects. A *post hoc* analysis of the hand arrangements using pairwise comparisons with Bonferroni correction showed significant differences between all three arrangements ($p < 0.01$). Users were 5.62% faster with the grid arrangement and 9.46% faster with the cube arrangement, compared against the single hand.

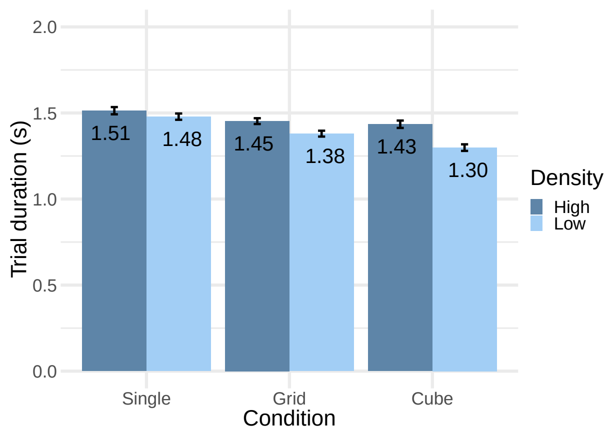

Figure 3: Mean trial completion time in seconds per hand arrangement and distractor density. Error bars represent 95% confidence intervals.

Trial completion times generally decreased with a higher amount of hands. This effect was larger in the low density than the high. This can be attributed to the queue algorithm we use for disambiguation, where an increase of hands will also lead to an increased time spent in the queue to cycle to the intended hand.

5.2 Error Rate and Disambiguation

An error is a trial in which the participant selected a distractor and not the target. The error rate is defined as the percentage of trials in which an error occurred. Table 2 shows the mean error rate. Participants had no errors in the one-handed arrangement, a mean error rate of 0.17% in the grid arrangement in high density and 0.04% in low density, as well as a mean error rate of 1.79% in the cube arrangement in high density and 0.25% in the low density.

The error rates are low across hand arrangements and densities. Interestingly, participants did not have a single error in the single-hand arrangement, suggesting a perfect understanding of the task, but something specific to the more-handed conditions causing trouble with the selection.

We observed that whenever errors did occur in the other hand arrangements, it tended to correlate with the activity of the queue algorithm that we use as a disambiguation technique. To demonstrate this, Table 2 also includes the overall percentage of trials in which the queue is active. To better understand how the queue affects the error rate, we further analysed how many error trials had an active queue in the grid and cube arrangements and the two distractor densities. The results of this can be seen in Figure 4.

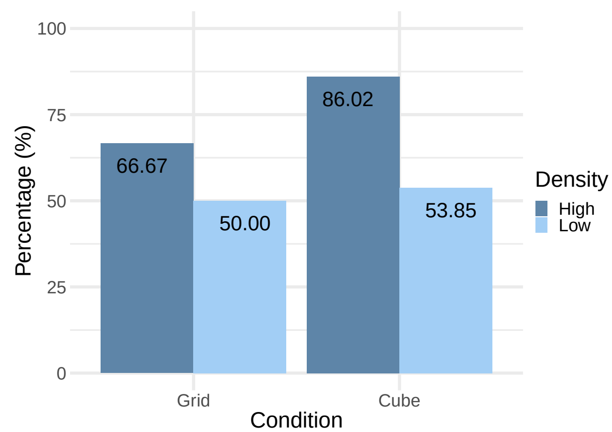

Figure 4: Percentage of errors during which the queue was active.

While the overall queue activity was quite low as seen in Table 2, Figure 4 shows that it was active during at least half of every error trial, with notably higher activity in the errors in the high density.

5.3 Motion

Motion is defined as the total distance in centimeters that the participant moved the physical controller within each successful trial. Figure 5 shows the mean motion per trial. The overall mean distances moved were 36.64cm for the single hand in high density and 35.88cm in low density, 25.74cm for the grid in high density and 24.62cm in low density, 17.40cm for the cube in high density and 15.47cm for low density.

Two-way repeated measures ANOVAs for motion, hand arrangement and distractor density showed a significant effect of hand arrangement ($F_{2,34} = 397.34, p < .001$) and distractor density ($F_{1,17} = 20.91, p < .001$). There was no interaction effect. A *post hoc* analysis of the hand arrangements using pair-wise comparisons with Bonferroni correction showed significant differences between all three arrangements ($p < 0.01$). Users moved 30.29% less with the grid arrangement and 53.94% less with the cube arrangement, compared against the single hand.

As hypothesized and per Fitts' law [14, 32], as well as prior research in adapting it to three-dimensional spaces [4, 46], participants generally moved less when the distance from the nearest hand to the intended target was lower.

	Single Mean	*Single* SD	*Grid* Mean	*Grid* SD	*Cube* Mean	*Cube* SD
Terrible (1) — Wonderful (7)	6.17	0.79	6.17	0.62	5.94	0.73
Frustrating (1) — Satisfying (7)	6.06	1.16	6.33	0.59	5.94	0.80
Dull (1) — Stimulating (7)	5.83	0.92	5.50	1.29	5.44	1.04
Difficult (1) — Easy (7)	6.17	1.25	6.33	1.03	6.28	0.89
Inadequate power (1) — Adequate power (7)	6.06	0.87	6.11	0.90	6.06	0.94
Rigid (1) — Flexible (7)	6.06	1.11	5.94	1.16	5.61	1.58
Smoothness during operation was: very rough (1) — very smooth (7)	6.17	0.92	6.11	0.83	6.28	0.67
The mental effort required for operation was: too low (1) — too high (7)	3.06	1.30	2.67	1.19	3.28	1.18
The physical effort required for operation was: too low (1) — too high (7)	3.17	1.20	3.00	1.37	2.94	1.16
Accurate pointing was: easy (1) — difficult (7)	2.28	1.41	2.17	1.29	2.28	1.27
General comfort was: very uncomfortable (1) — very comfortable (7)	6.17	0.86	6.00	1.08	6.22	0.73

Table 1: Subjective satisfaction of participants for each hand arrangement. Questionnaire from Hornbæk & Hertzum [23], normalized to a 7-point scale.

Arrangement	*High* *Single*	*High* *Grid*	*High* *Cube*	*Low* *Single*	*Low* *Grid*	*Low* *Cube*
Error Rate (%)	0.0	0.17	1.79	0.0	0.04	0.25
Queue Active (%)	0.0	1.87	7.33	0.0	0.41	2.08

Table 2: Mean error rates and overall queue activity in percentage across hand arrangements and distractor densities.

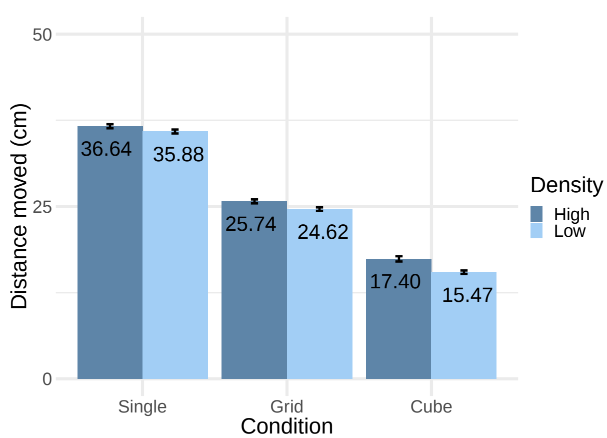

Figure 5: Mean trial motion per hand arrangement and distractor density. Error bars represent 95% confidence intervals.

Surprisingly, this effect was present even in the single-hand condition and therefore cannot be attributed entirely to the queue. We observed that users were prone to moving the hand around targets during the single-hand condition specifically, which explains this.

5.4 Subjective Satisfaction

Table 1 shows the data collected from the subjective satisfaction questionnaires administered following each hand arrangement. We used the aligned rank transform for non-parametric factorial analyses by Wobbrock et al. [49] to evaluate the effects of the on the TLX measures. This revealed no significant effects, so we can only report on trends. Looking at the means, for the first six questions, the *Ninja Hands* arrangements generally outperform the single hand. Smoothness sees the grid arrangement score slightly lower than the single hand. Mental and physical effort trends towards the ideal center for those scores (middle, or 3.5). However, mental effort for the grid condition ranks notably (>10%) lower than the other two.

6 SECOND STUDY

Our initial study showed that as the number of hands increased, the motion exerted and time taken to select a target decreased. While those results demonstrate performance improvements for *ninja hands*, it is unclear if these benefits can be retained with higher numbers of hands. While the first study investigated performance in a relatively small target space, we want to better understand of how performance translates to larger distances, and larger arrangements of targets. We implement a follow-up study that seeks to determine the impact of these factors.

6.1 Design

We adapt the target selection task from the first study, taking place in the same 10m×5m×10m environment with targets filling the room (Figure 6). We manipulate the *hand count*, and control the distance to the targets.

We extend the cubic hand arrangement, evaluating *hand count* by comparing one hand against cubes of 2×2×2 (8), 3×3×3 (27), and 4×4×4 (64) *Ninja Hands*. The hands are distributed across the room so that they fill out the space regardless of the number of hands; this is done by subdividing the space into a number of cuboids equal to the number of hands and placing each hand in the center of such a cuboid. This is the three-dimensional equivalent of the pattern used in *ninja cursors* [28]. This pattern causes the eight hands to

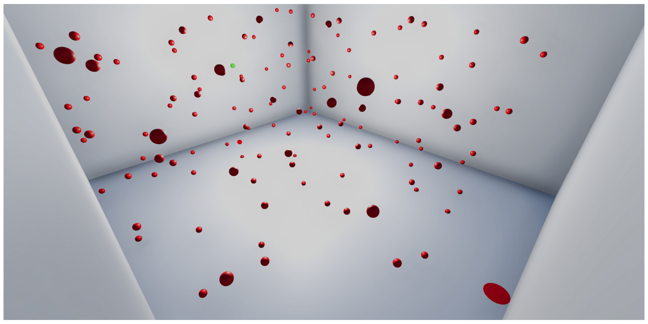

Figure 6: The first of eight target and distractor layouts in the second study, seen from the top corner of the room. Participants stand on the red mark on the floor at the back of the room.

be distributed 5m apart, the 27 hands are 3.33m apart, and the 64 hands are 2.5m apart. All hands have a fixed movement gain of 10x across all conditions. This gain was decided based on a criterion that the single hand centered in the room can comfortably reach any target in the room.

The targets are placed at one of four distances from the center of the room in a random direction. These distances are 1.5m, 3m, 4.5m, and 6m; this corresponds to 20%, 40%, 60%, and 80% of the half-diagonal of the room (7.5m). We generate three targets at each distance, discarding those that fall outside the bounds of the room. Then, we randomize their selection order. This process is repeated eight times for $8 \times 12 = 96$ individual targets, and is fixed between participants, to minimize the effects of any one configuration. As in the first study, we populate the rest of the room with distractors using Poisson disk sampling [10] to simulate a busy target space, however only at one density; 10% of the diagonal of the room (1.5m). This approach to target generation ensures a variety of depths and angular sizes for hands and targets, and ensures a greater variety of these factors than in the first study.

Each participant selects all targets in a fixed order using the different hand counts in the four conditions. We balance the order of the conditions between participants using a Latin square. Altogether, this gives us 384 observations (96 targets $\times \times$ 4 hand counts) per participant.

6.2 Measures

As in the first study, we report the completion time for each trial, movement of the physical controllers, and error rates.

We use NASA TLX [19] to measure how subjective workload is affected by different hand counts. This replaces the user satisfaction questionnaire [23] from the first study, since the only score that differed in that data was mental load for one of the more-handed conditions for one of the more-handed conditions (see Table 1). NASA TLX is one of the most common questionnaires for measuring subjective workload [18, 19].

6.3 Apparatus

The apparatus is as in the first study.

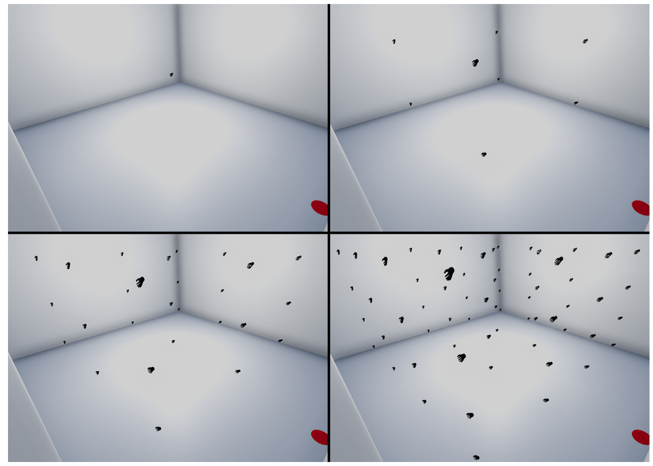

Figure 7: The four numbers and arrangements of hands used in the second study, centered in the room and seen from the top corner. From top left: 1, 8, 27, and 64 hands. The arrangement is the three-dimensional equivalent of the even arrangement used in *ninja cursors* [28], adapted to a non-cubic environment. The user stands on the red mark on the floor.

6.4 Participants

We recruited 20 participants (9 female), age range 21-48, mean = 28.6, SD = 6.2. We followed a similar recruitment procedure as in the first study; participants must not have participated in that study.

6.5 Procedure

The procedure is similar to the first study. However, there are three changes to accommodate the new design.

First, between each condition, participants fill out the NASA TLX questionnaire [18, 19] instead of the user satisfaction questionnaire [23].

Second, inside the virtual environment, the participant is placed on a red marker on the floor in the back center of the room, moved back a bit from the target space to ensure that as many hands as possible are visible when looking forward, even in the many-handed conditions.

Third, as in the first study, participants have a reset step between each trial. The hands are shown between each step, to minimize time spent finding the hands.

7 SECOND STUDY RESULTS

This section describes the results of the second study. We first discard the first round of 12 trials for each condition to balance initial training effects (12.5% of the total number of trials). We then discard outliers, defined as data points that fall outside 1.5 times the interquartile range for completion time and motion. 1.1% outliers were discarded for completion time and 0.6% outliers were discarded for motion. We take these to represent trials where some external factor, such as momentary loss of tracking, caused participants to idle or move far more than they otherwise would.

7.1 Completion Time

Completion time is calculated as the difference between the moment when the participant ends the resting phase by pressing the trigger and the moment when the participant successfully selects the target by pressing the trigger when a hand is touching it.

The raw mean completion times were 3.59 seconds for the 1-handed condition, 3.66 seconds for the 8-handed condition, 3.61 seconds for the 27-handed condition, and 3.69 seconds for the 64-handed condition. Since we seek to understand how much of this time was actually spent moving, we isolate the time spent deciding which hand to use. To accomplish this, we create a filter that for each trial logs how long it takes for the participant to move the controller more than 10 cm from where they started. We make the assumption that participants do not move much until they have decided how to move.

Isolating the decision-making time reveals a linear increase in decision-making time as we increase hand count, as seen in Figure 8. The mean decision time is 1.38 seconds in the 1-handed condition, 1.67 seconds in the 8-handed condition, 1.97 seconds in the 27-handed condition, and 2.26 seconds in the 64-handed condition.

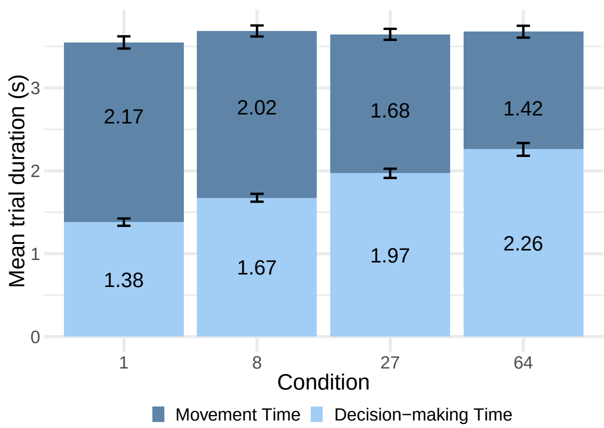

Figure 8: Mean trial duration for the second study in seconds, broken down into decision-making time and movement time. Error bars represent 95% confidence intervals for decision-making time and movement time.

We reason that the linear relationship between hand count and decision-making time may be explained by the number of hands adjacent to the optimal hand increasing by a like amount when using our arrangement, regardless of the hand count.

Adjusting the overall trial duration by each trial's decision-making time tells us how much time was spent moving. This reveals a significant reduction in time spent moving as hand count increases, as shown in Figure 8 and isolated in Figure 9. The movement time is 2.17 seconds in the one-handed condition, 2.02 seconds in the 8-handed condition, 1.68 seconds in the 27-handed condition, and 1.42 seconds in the 64-handed condition.

We performed one-way repeated measures ANOVAs for overall trial duration, decision-making time, and movement time. This revealed no significant effect of hand count on overall trial duration ($F_{1,3} = 1.70, p = .21$), but a significant effect of hand count

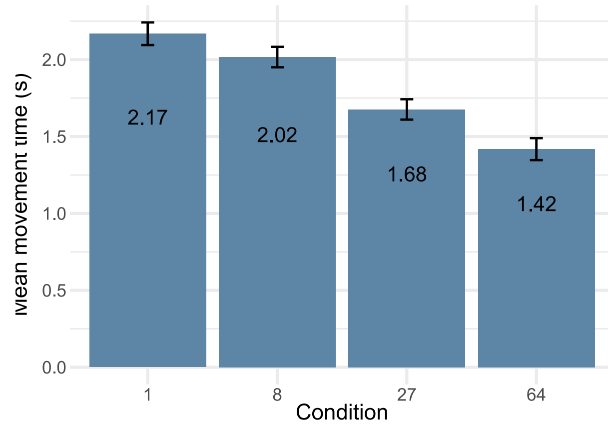

Figure 9: Mean movement time per condition in the second study. Error bars represent 95% confidence intervals.

on decision-making time ($F_{1,3} = 68.08, p < .01$) and movement time ($F_{1,3} = 34.52, p < .01$). This demonstrates that while overall trial durations are comparable across conditions, there is an expected trade-off between time spent deciding which hand to use and reduces physical motion exerted.

7.2 Error Rate and Disambiguation

An error is a trial in which the participant selected a distractor and not the target. The error rate is defined as the percentage of trials in which an error occurred. Table 3 shows the mean error rate and queue activity in the second study. We see a similar pattern in error rate as in the first study. Participants had no errors in the one-handed arrangement, a mean error rate of 0.05% with 8 hands, 0.23% with 27, and 0.48% with 64. Like the first study, this suggests that participants understood the task, but something in the more-handed conditions caused the occasional wrongful selection.

	Condition			
	1	*8*	*27*	*64*
Error Rate (%)	0.0	0.05	0.23	0.48
Queue Active (%)	0.0	0.38	1.81	3.95

Table 3: Mean error rates and overall queue activity in percentage across the four hand conditions in the second study.

We again see similar correlations between queue activity and error rates, though overall error rates are very low.

Figure 10 shows the percentage of error trials during which the queue was active. Curiously, the eight-handed condition has a higher activity than the other two more-handed conditions, though this can likely be explained by a very small sample size (0.05% of trials in the 8-handed condition were errors), which allows for large variation in this measure with just a few trials. Overall, these data indicate that disambiguation did not contribute much to target selection time, even with 64 hands and many targets arranged in the room per the configuration in the study.

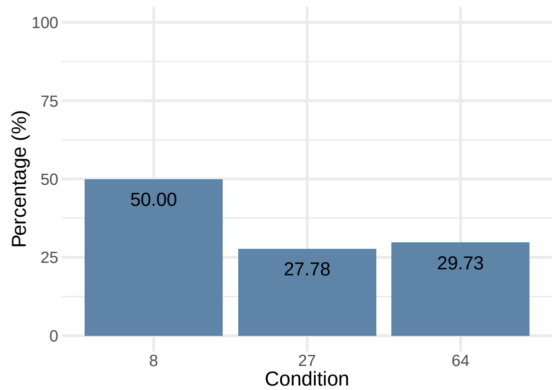

Figure 10: Percentage of error trials during which the queue was active in the second study.

7.3 Motion

Motion is defined as the total distance in centimeters that the participant moved the physical controller within each successful trial. Figure 11 shows the mean distance moved per trial for the four conditions, as well as the optimal raw distance, defined as the lowest possible motion required to reach the target, not accounting for any queue activity that may increase motion distance. Participants moved 74cm on average in the 1-handed condition, 60cm in the 8-handed condition, 46cm in the 27-handed condition, and 40cm in the 64-handed condition.

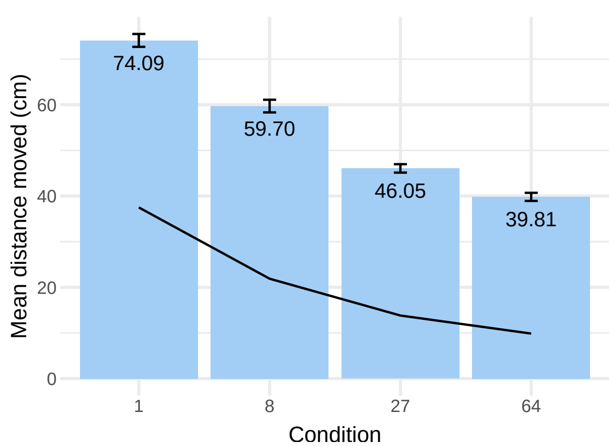

Figure 11: Mean distance moved per condition. The black line represents the optimal raw distance. Error bars represent 95% confidence intervals.

A one-way repeated measures ANOVA showed a significant effect of hand count on motion ($F_{1,3} = 19.48, p < .02$). As expected, participants moved significantly less with an increased hand count, showing that one of the main value propositions of *Ninja Hands* holds even at higher hand counts. Participants do move more than the optimal distance; however, the decision-making time is included in these data (which by extension comes any initial motion while locating the target), as is any potential queue activity on the way. Since we apply a universal scaling factor of 10 times to the mapping to keep conditions as comparable as possible, this also leads to some universal loss of precision; one of the critical benefits of *Ninja Hands* is the ability to reduce this scaling factor and still reach distant targets.

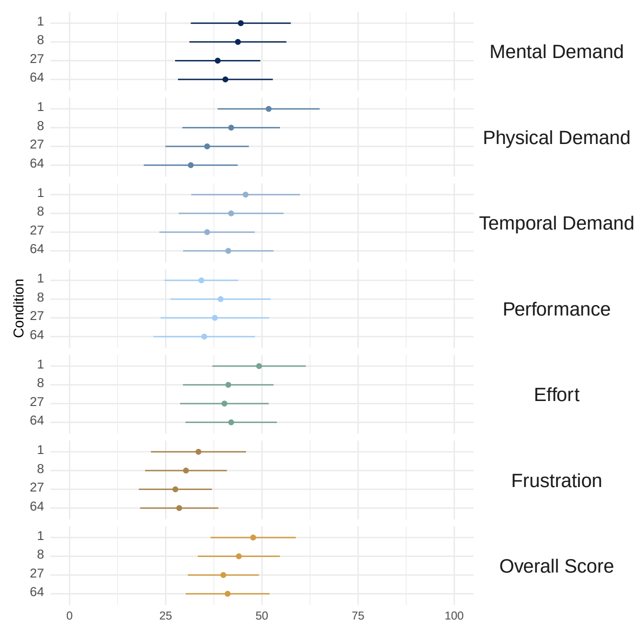

Figure 12: Mean NASA TLX results per scale and overall weighted scores, for each condition. Error bars represent 95% confidence intervals.

7.4 NASA TLX

Table 4 and Figure 12 show the results of the NASA TLX. We used the aligned rank transform for non-parametric factorial analyses by Wobbrock et al. [49] to evaluate the effects of the hand conditions on the TLX measures. The only significant effect was on physical demand ($F_{1,3} = 7.21, p < .001$). This corroborates our motion data. For the other measures, there are only general, but insignificant, trends. For our main finding, as hypothesized, overall workload decreases from the 1-handed (47.71) to the 8-handed (44) and 27-handed (39.96) conditions, following a slight increase for the 64-handed (41.08) condition. A similar trend of a reduction followed by a slight increase for the 64-handed condition can be seen individually in effort, frustration, temporal demand, and mental demand. Curiously, participants feel more performant in the 1-handed (34.25) condition than the 8-handed (39.25) and 27-handed (37.75) conditions, though the 64-handed condition (35) sees a slight gain in this measure.

Altogether, these data demonstrate that though more-handed conditions trend toward a lower workload, this gain seems to even out and potentially reverse at higher hand counts due to the experiences of effort, frustration and mental demand that can be linked to the increased number of hands to choose from and the increase in time spent making this decision (Figure 8). However, the only significant effect is for the experience of physical demand, which correlates with the actual physical motion exerted (Figure 11).

8 DISCUSSION

We have described the *Ninja Hands* technique, evaluated it through two studies, and shown that it can achieve a significant reduction in the physical motion exerted. However, we only found significant gains in target selection speed in a smaller environment in front of the user. In a larger environment with more hands, we found a significant decrease in time spent moving with more hands, resulting in a corresponding increase in time spent deciding which hand to use. Thereby, overall selection speeds were comparable across hand counts. Lastly, while the subjective workload was lower with more hands, the results indicate that this benefit reverses with the increase from 27 to 64 hands. More hands come at a cost. This section discusses how these findings might be applied to future implementations of the technique.

8.1 Improving Target Selection

Our results show that having a smaller number of hands in a smaller environment, such as in the first study, can reduce target selection speed. Further, the results indicate that extending this to a larger environment with more hands, as in the second study, results in stable target selection speed without significant difference across conditions. If the goal is to reduce target selection speed, using a small number of hands in a smaller environment is optimal. An exponential (n^3) increase in the number of hands only causes a linear increase in decision-making time. This relation is likely due to the fact that the number of hands immediately surrounding the optimal hand is not increasing exponentially, even if the overall number of hands in the environment does. Thus, while simply filling the environment with hands does not improve selection speed, using multiple smaller clusters of hands in the scene might be beneficial.

For example, imagine a typical VR scene such as a living room, where objects are spread across tables, couches, and shelves (Figure 1). A specialized hand arrangement can be generated based on the layout of the objects; the objects can be grouped using a clustering algorithm and then based on the radius of these clusters, a number of hands can be generated so that the user can reach all objects comfortably. Further, this number can be optimized, so that little to no scaling factor is needed; participants in the second study moved more than they had to in all conditions, suggesting that having a lower scaling factor is ideal. The ability to lower the scaling factor, while still allowing all objects to be easily reached, is a key benefit of *Ninja Hands*. However, in a typical VR scenario, the layout of items will change. When selecting an object, for instance, users will typically move it, modifying the cluster. Of course, the scaling factor can be modified dynamically by recalculating the clusters, but having the mapping of hands change on the fly might make controlling the hands difficult.

The results indicate that when the scaling factor is constant across hand counts, higher hand counts still select targets faster than fewer hands in a smaller environment. Further, higher scaling factors appear to lead to lower precision. A stable, low scaling factor appears ideal. Therefore, an alternate approach would change the hand count instead of the scaling factor, so that hands are dynamically added or removed to each cluster as the configuration of objects changes. We envision an implementation of *Ninja Hands* wherein these clusters of smaller hand counts are generated based on the object layout in the environment. If the clusters change due to manipulation of the objects, the hand count is dynamically updated to ensure that all targets can be comfortably reached.

8.2 Improving Existing Techniques

Ninja Hands can be integrated with existing selection techniques and offers other improvements than just target selection speed in smaller environments. A reduction in the physical motion exerted can be reliably obtained by increasing hand count based on the arrangements used in our studies, and overall lower subjective workloads in target selection can be gained, which might benefit other techniques. For example, we envision an implementation where the user can activate the *Ninja Hands* with the press of a button, which multiplies their hand and moves the hands into an even arrangement in the environment to achieve reduced motion exertion and workload. Then, once the user has made their selection, the object is moved back to the user's perspective in their hand; alternatively, it could be integrated with teleportation, so that the user is teleported adjacent to the objected selected using *Ninja Hands*.

8.3 Improving Disambiguation

In our studies, disambiguation played a limited role. Given the arrangements and numbers of hands and targets used in the two studies, the queue algorithm was rarely used in the many-hand conditions. Similar work with cursors in desktop environments, such as *ninja cursors* [28], use disambiguation more frequently. It is a limitation of our study designs that limited disambiguation happened. Different ratios between the placement and density of hand and target arrangements might cause more problems with disambiguation. For example, our second study shows that *Ninja Hands* can perform in large virtual environments, which is an improvement over existing hand-based techniques. However, had the scene in that study contained very dense clusters of targets, more hands would enter the queue and the queue would take longer to resolve.

Recent work comparing how selection techniques perform in high-density conditions [30] has demonstrated how selection mechanisms similar to *bubble cursor* [16, 48] can be integrated with ray-based selection to function well in scenes with high target density. However, even if the ray is used to move a hand, this method of pointing to control the hand is quite different to using natural hand motion to reach targets, which is a potential advantage of the *Ninja Hands* technique. We envision an integration of these two concepts that retains their individual benefits. VR environments sometimes contain clusters of objects spaced apart, such as tables and shelves, each containing numerous objects that the user may interact with. Naively filling such an environment with hands would cause issues with occlusion, distraction, and disambiguation. Instead, one arrangement of hands could be placed at each cluster of objects, which should allow each cluster to retain similar benefits in motion

	Condition							
	1		*8*		*27*		*64*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Effort	49.20	25.96	41.25	25.17	40.25	24.62	42.00	25.41
Frustration	33.50	26.41	30.25	22.73	27.50	20.29	28.50	21.70
Mental Demand	44.50	27.76	43.75	26.94	38.50	23.68	40.50	26.35
Performance	34.25	20.47	39.25	20.81	37.75	30.19	35.00	28.19
Physical Demand	51.75	28.34	42.00	27.11	35.75	23.18	31.50	26.11
Temporal Demand	45.75	30.18	42.00	29.17	35.75	26.47	41.25	25.12
Overall Score	**47.71**	23.68	**44.00**	22.80	**39.96**	19.72	**41.08**	23.32

Table 4: NASA TLX results per scale and overall weighted scores, for each condition.

exerted and selection speed as in our first study. The issue, then, becomes how to disambiguate between clusters of hands, so that only one cluster is active at a time. For this purpose, a ray-based heuristic could be used, such as the gaze-filtered bubble ray as suggested in [30], as our decision-making results indicate that users take the time to look for the hand they want to use before attempting to use it.

8.4 Limitations

There are three key limitations to our work. First, a well-established benefit of using hands in VR is their positive effects on aspects of embodiment, often discussed as the sum of body ownership, self-location, and agency [25, 29]. We do not investigate embodiment in this work. Ownership of a full body has been achieved from a third-person perspective given synchronous visuo-tactile stimulation [7], and this has even been extended to two bodies [21]. The sense of self-location is also malleable and can be consistently manipulated from a third-person perspective [33], which also extends to two bodies [21]. However, it is unclear if these findings apply to such great distances, numbers, and floating hands as investigated in our work. For future work, we suggest investigating how the synchronous use of many hands that are distantly removed from the user's perspective affects the sense of ownership, whether this sense can be modulated by synchronous visuo-tactile stimulation, and whether the sense of self-location shifts towards hands that are more often used. Further, the sense of presence could be affected by the number of hands as compared to ordinary VR experiences, and this could also be investigated.

Second, the primary purpose of selecting objects in VR with hands is to then manipulate them, typically by moving them elsewhere. Our work focuses on selection, rather than manipulation. It would be interesting to explore how the *Ninja Hands* concepts could improve manipulation. For example, we envision being able to pass objects from one hand to adjacent hands based on directional motion, swiftly moving objects across large distances. Further, we suggest exploring the use of many hands as a metaphor for weight for very large objects: Having to use dozens of hands to manipulate larger objects could be a good way to communicate how heavy they are, while also limiting accidental, jerky motion.

Third, we only investigated the technique from the perspective of uni-manual interaction; participants in our studies always controlled the many hands with one physical hand. While we see no reason to expect that the technique would not translate to bi-manual interaction, it could be interesting to investigate specific interactions enabled by using two physical hands. It could perhaps be an intuitive way to apply variable mapping to the hands, so that one physical hand controls distant virtual hands with a non-linear mapping function, while the other physical hands controls closer virtual hands with a constant mapping.

9 CONCLUSION

We investigate the concept of using many distributed hands in VR to improve target selection at a distance. We introduce *Ninja Hands*, an interaction technique that maps one physical hand to many distributed virtual hands. We examine the technique in two empirical studies and find significant reductions in motion exerted, as well as conditional reductions in target selection time. We also find a trade-off between decision-making time and time spent moving, as well as a benefit of reduced workload that our data suggest reverses at higher hand counts. We discuss how these findings can contribute towards future versions of the *Ninja Hands* technique and can help improve target selection in VR. Altogether, our work introduces many hands for target selection in VR and outlines benefits and drawbacks associated with this concept.

ACKNOWLEDGMENTS

We want to thank Valve Corporation for providing the prototype controllers used in our experiments. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 648785).

REFERENCES

[1] F. Argelaguet and C. Andujar. 2009. Efficient 3D Pointing Selection in Cluttered Virtual Environments. *IEEE Computer Graphics and Applications* 29, 6 (Nov 2009), 34–43. https://doi.org/10.1109/MCG.2009.117

[2] Ferran Argelaguet, Ludovic Hoyet, Michaël Trico, and Anatole Lécuyer. 2016. The role of interaction in virtual embodiment: Effects of the virtual hand representation. *Proceedings - IEEE Virtual Reality* 2016-July (2016), 3–10. https://doi.org/10.1109/VR.2016.7504682

[3] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2016. Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 1968–1979. https://doi.org/10.1145/2858036.2858226

[4] Ravin Balakrishnan. 2004. "Beating" Fitts' Law: Virtual Enhancements for Pointing Facilitation. *Int. J. Hum.-Comput. Stud.* 61, 6 (Dec. 2004), 857–874. https://doi.org/10.1016/j.ijhcs.2004.09.002

[5] Joanna Bergström, Aske Mottelson, Andreea Muresan, and Kasper Hornbæk. 2019. Tool Extension in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 568, 11 pages. https://doi.org/10.1145/3290605.3300798

[6] Renaud Blanch and Michaël Ortega. 2009. Rake Cursor: Improving Pointing Performance with Concurrent Input Channels. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. ACM, New York, NY, USA, 1415–1418. https://doi.org/10.1145/1518701.1518914

[7] Pierre Bourdin, Itxaso Barberia, Ramon Oliva, and Mel Slater. 2017. A Virtual Out-of-Body Experience Reduces Fear of Death. *PLOS ONE* 12, 1 (jan 2017), e0169343. https://doi.org/10.1371/journal.pone.0169343

[8] Doug A. Bowman and Larry F. Hodges. 1997. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics* (Providence, Rhode Island, USA) *(I3D '97)*. ACM, New York, NY, USA, 35–ff. https://doi.org/10.1145/253284.253301

[9] Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. 2016. Point & and Teleport Locomotion Technique for Virtual Reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (Austin, Texas, USA) *(CHI PLAY '16)*. ACM, New York, NY, USA, 205–216. https://doi.org/10.1145/2967934.2968105

[10] 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

[11] Han Joo Chae, Jeong-in Hwang, and Jinwook Seo. 2018. Wall-based Space Manipulation Technique for Efficient Placement of Distant Objects in Augmented Reality. In *The 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18*. ACM Press, New York, New York, USA, 45–52. https://doi.org/10.1145/3242587.3242631

[12] Tiare Feuchtner and Jörg Müller. 2017. Extending the Body for Interaction with Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 5145–5157. https://doi.org/10.1145/3025453.3025689

[13] Tiare Feuchtner and Jörg Müller. 2018. Ownershift: Facilitating Overhead Interaction in Virtual Reality with an Ownership-Preserving Hand Space Shift. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. ACM, New York, NY, USA, 31–43. https://doi.org/10.1145/3242587.3242594

[14] Paul M. Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47, 6 (1954), 381–391. https://doi.org/10.1037/h0055392

[15] Andrew Forsberg, Kenneth Herndon, and Robert Zeleznik. 1996. Aperture based selection for immersive virtual environments. In *Proceedings of the 9th annual ACM symposium on User interface software and technology - UIST '96*. ACM Press, New York, New York, USA, 95–96. https://doi.org/10.1145/237091.237105

[16] Tovi Grossman and Ravin Balakrishnan. 2005. The Bubble Cursor: Enhancing Target Acquisition by Dynamic Resizing of the Cursor's Activation Area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. ACM, New York, NY, USA, 281–290. https://doi.org/10.1145/1054972.1055012

[17] Arvid Guterstam, Valeria I. Petkova, and H. Henrik Ehrsson. 2011. The Illusion of Owning a Third Arm. *PLoS ONE* 6, 2 (feb 2011), e17208. https://doi.org/10.1371/journal.pone.0017208

[18] 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 (oct 2006), 904–908. https://doi.org/10.1177/154193120605000909

[19] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *SpringerReference*. Springer-Verlag, Berlin/Heidelberg, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[20] Kenneth P. Herndon, Andries van Dam, and Michael Gleicher. 1994. The Challenges of 3D Interaction: A CHI '94 Workshop. *SIGCHI Bull.* 26, 4 (Oct. 1994), 36–43. https://doi.org/10.1145/191642.191652

[21] Lukas Heydrich, Trevor J. Dodds, Jane E. Aspell, Bruno Herbelin, Heinrich H. Bülthoff, Betty J. Mohler, and Olaf Blanke. 2013. Visual capture and the experience of having two bodies – Evidence from two different virtual reality techniques. *Frontiers in Psychology* 4, DEC (2013), 1–15. https://doi.org/10.3389/fpsyg.2013.00946

[22] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. 1994. A survey of design issues in spatial input. In *Proceedings of the 7th annual ACM symposium on User interface software and technology - UIST '94*. ACM Press, New York, New York, USA, 213–222. https://doi.org/10.1145/192426.192501

[23] Kasper Hornbæk and Morten Hertzum. 2007. Untangling the usability of fisheye menus. *ACM Transactions on Computer-Human Interaction* 14, 2 (2007), 6–es. https://doi.org/10.1145/1275511.1275512

[24] Ludovic Hoyet, Ferran Argelaguet, Corentin Nicole, and Anatole Lécuyer. 2016. "Wow! I Have Six Fingers!": Would You Accept Structural Changes of Your Hand in VR? *Frontiers in Robotics and AI* 3, MAY (may 2016), 1–12. https://doi.org/10.3389/frobt.2016.00027

[25] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2013. The Sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments* 22, 1 (2013), 373–387.

[26] Konstantina Kilteni, Jean-Marie Normand, Maria V. Sanchez-Vives, and Mel Slater. 2012. Extending Body Space in Immersive Virtual Reality: A Very Long Arm Illusion. *PLOS ONE* 7, 7 (07 2012), 1–15. https://doi.org/10.1371/journal.pone.0040867

[27] Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 345, 9 pages. https://doi.org/10.1145/3173574.3173919

[28] Masatomo Kobayashi and Takeo Igarashi. 2008. Ninja Cursors: Using Multiple Cursors to Assist Target Acquisition on Large Screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, NY, USA, 949–958. https://doi.org/10.1145/1357054.1357201

[29] Matthew R. Longo, Friederike Schüür, Marjolein P.M. Kammers, Manos Tsakiris, and Patrick Haggard. 2008. What is embodiment? A psychometric approach. *Cognition* 107, 3 (jun 2008), 978–998. https://doi.org/10.1016/j.cognition.2007.12.004

[30] Y. Lu, C. Yu, and Y. 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

[31] Paul Lubos, Gerd Bruder, and Frank Steinicke. 2014. Are Four Hands Better Than Two?: Bimanual Interaction for Quadmanual User Interfaces. In *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction* (Honolulu, Hawaii, USA) *(SUI '14)*. ACM, New York, NY, USA, 123–126. https://doi.org/10.1145/2659766.2659782

[32] I. Scott MacKenzie and William Buxton. 1992. Extending Fitts' Law to Two-dimensional Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. ACM, New York, NY, USA, 219–226. https://doi.org/10.1145/142750.142794

[33] Antonella Maselli and Mel Slater. 2014. Sliding perspectives: Dissociating ownership from self-location during full body illusions in virtual reality. *Frontiers in Human Neuroscience* 8, September (2014), 1–19. https://doi.org/10.3389/fnhum.2014.00693

[34] Sven Mayer, Valentin Schwind, Robin Schweigert, and Niels Henze. 2018. The Effect of Offset Correction and Cursor on Mid-Air Pointing in Real and Virtual Environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 653, 13 pages. https://doi.org/10.1145/3173574.3174227

[35] Jeffrey S. Pierce, Brian C. Stearns, and Randy Pausch. 1999. Voodoo Dolls: Seamless Interaction at Multiple Scales in Virtual Environments. In *Proceedings of the 1999 Symposium on Interactive 3D Graphics* (Atlanta, Georgia, USA) *(I3D '99)*. ACM, New York, NY, USA, 141–145. https://doi.org/10.1145/300523.300540

[36] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. The Go-go Interaction Technique: Non-linear Mapping for Direct Manipulation in VR. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology* (Seattle, Washington, USA) *(UIST '96)*. ACM, New York, NY, USA, 79–80. https://doi.org/10.1145/237091.237102

[37] Ivan Poupyrev and Tadao Ichikawa. 1999. Manipulating Objects in Virtual Worlds: Categorization and Empirical Evaluation of Interaction Techniques. *Journal of Visual Languages and Computing* 10, 1 (1999), 19 – 35. https://doi.org/10.1006/jvlc.1998.0112

[38] Kari-Jouko Räihä and Oleg Špakov. 2009. Disambiguating Ninja Cursors with Eye Gaze. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. ACM, New York, NY, USA, 1411–1414. https://doi.org/10.1145/1518701.1518913

[39] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017. "These Are Not My Hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 1577–1582. https://doi.org/10.1145/3025453.3025602

[40] Valentin Schwind, Jan Leusmann, and Niels Henze. 2019. Understanding Visual-Haptic Integration of Avatar Hands Using a Fitts' Law Task in Virtual Reality. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 211–222. https://doi.org/10.1145/3340764.3340769

[41] Valentin Schwind, Sven Mayer, Alexandre Comeau-Vermeersch, Robin Schweigert, and Niels Henze. 2018. Up to the Finger Tip: The Effect of Avatars on Mid-Air Pointing Accuracy in Virtual Reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) *(CHI PLAY '18)*. ACM, New York, NY, USA, 477–488. https://doi.org/10.1145/3242671.3242675

[42] Ben Shneiderman and Catherine Plaisant. 2004. *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley.

[43] Mel Slater, Bernhard Spanlang, Maria V. Sanchez-Vives, and Olaf Blanke. 2010. First Person Experience of Body Transfer in Virtual Reality. *PLOS ONE* 5, 5 (05 2010), 1–9. https://doi.org/10.1371/journal.pone.0010564

[44] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of Presence in Virtual Environments. *Presence: Teleoper. Virtual Environ.* 3, 2 (Jan. 1994), 130–144. https://doi.org/10.1162/pres.1994.3.2.130

[45] Richard Stoakley, Matthew J Conway, and Randy Pausch. 1995. Virtual reality on a WIM. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95*. ACM Press, New York, New York, USA, 265–272. https://doi.org/10.1145/223904.223938

[46] Wolfgang Stuerzlinger and Robert J. Teather. 2014. Considerations for Targets in 3D Pointing Experiments. In *Proceedings of HCI Korea* (Seoul, Republic of Korea) *(HCIK '15)*. Hanbit Media, Inc., South Korea, 162–168. http://dl.acm.org/citation.cfm?id=2729485.2729510

[47] Huawei Tu, Susu Huang, Jiabin Yuan, Xiangshi Ren, and Feng Tian. 2019. Crossing-Based Selection with Virtual Reality Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 618, 14 pages. https://doi.org/10.1145/3290605.3300848

[48] L. Vanacken, T. Grossman, and K. 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. https://doi.org/10.1109/3DUI.2007.340783

[49] 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 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, New York, New York, USA, 143. https://doi.org/10.1145/1978942.1978963

[50] H. P. Wyss, R. Blach, and M. Bues. 2006. iSith - Intersection-based Spatial Interaction for Two Hands. In *3D User Interfaces (3DUI'06)*. IEEE, Alexandria, VA, USA, 59–61. https://doi.org/10.1109/VR.2006.93

[51] Shumin Zhai, William Buxton, and Paul Milgram. 1994. The Silk Cursor: Investigating Transparency for 3D Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. ACM, New York, NY, USA, 459–464. https://doi.org/10.1145/191666.191822

[52] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. ACM, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053