ShadowClones: an Interface to Maintain a Multiple Sense of Body-space Coordination in Multiple Visual Perspectives

Kazuma Takada
Sony Computer Science Laboratories, Inc./Okinawa Institute of Science and Technology Graduate University

Nanako Kumasaki
Sony Computer Science Laboratories, Inc. / University of Tsukuba

Tom Froese
Okinawa Institute of Science and Technology Graduate University

Kazuhisa Shibata
Center for Brain Science, RIKEN

Jun Nishida
Sony Computer Science Laboratories, Inc. / University of Maryland, College Park

Shunichi Kasahara
Sony Computer Science Laboratories, Inc./Okinawa Institute of Science and Technology Graduate University

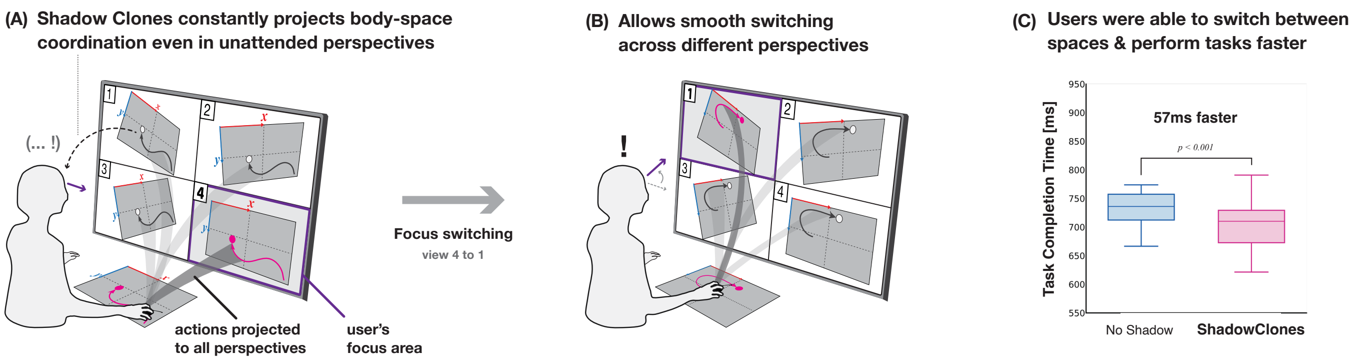

Figure 1: (A) We propose Shadow Clones interface that supports parallel interactions such as controlling multiple avatars or robots simultaneously. It projects a user's body movements to all perspectives, even in unattended areas, providing a sense of body-space coordination in the peripheral vision field. (B) This achieves seamless transitions from one to another workspace (57ms faster when compared to no assistance, as shown in C) thanks to a lower cognitive load for spatial recognition.

ABSTRACT

In this paper, we propose *ShadowClones*, an interface that supports interactions in which a single user can interact with multiple bodies in multiple spaces. Recent teleoperation technologies have allowed a user controlling multiple objects simultaneously, but at the same time, it also exhibited a significant challenge, which can be attributed to the high cognitive load caused by switching and recogning various spaces/perspectives repeatedly and instantly. To tackle this challenge, by taking advantage of pre-attentive visual cues for users' simultaneous information processing, we designed and evaluated a new user interface, called Shadow Clones, that projects self-body information in unattended areas for increasing the awareness of body-space relationships and allowing users to seamlessly switch across different visual perspectives from avatars or remote robots. We then explored the proposed approach through a simple visual reaching task with a performance evaluation in terms of task completion time and success rate. The results showed superior performance when compared with a condition that presents no projections of users' body movements in unattended areas. We conclude by discussing possible mechanisms of this enhancement as well as two potential scenarios using the shadow clones approach, including new entertainment content for virtual reality e-sports and multiple robot teleoperation such as in a construction site or a disaster site, without compromising operational performance.

CCS CONCEPTS

• **Human-centered computing → User interface design**.

KEYWORDS

multiple spatial perception, multiple body projection, motor control, unattended visual processing, parallel interaction

ACM Reference Format:

Kazuma Takada, Nanako Kumasaki, Tom Froese, Kazuhisa Shibata, Jun Nishida, and Shunichi Kasahara. 2023. ShadowClones: an Interface to Maintain a Multiple Sense of Body-space Coordination in Multiple Visual Perspectives. In *Augmented Humans Conference (AHs '23), March 12–14, 2023, Glasgow, United Kingdom.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3582700.3582706

AHs '23, March 12–14, 2023, Glasgow, United Kingdom

ACM ISBN 978-1-4503-9984-5/23/03…$15.00
https://doi.org/10.1145/3582700.3582706

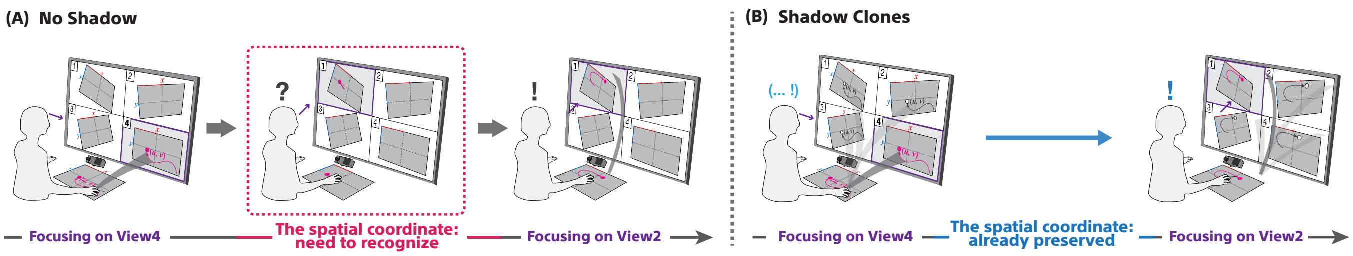

Figure 2: (A) No Shadow interface: the cursor aphpears only in a space where the user is gazing at. (B) Our ShadowClones interface: the user's cursor is presented in all spaces. Showing an active cursor (in red color) in a focused area as well as shadow cursors (in white color) can preserve the sense of body-space coordination in multiple visual spaces.

1 INTRODUCTION

The development of real-time network communication technologies has enabled the simultaneous operation of multiple virtual avatars [31] or remote robots by a single user from a remote location (e.g., controlling remote construction machines [12, 26] or playing a team sport by a single user [32, 43]). These studies have demonstrated the feasibility of a novel interaction style in which a single user can interact in multiple spaces simultaneously.

Furthermore, although the potential of a parallel interaction style that is beyond a user's physical capability has been proven, a significant challenge still exists, which can be attributed to the **high cognitive load caused by switching and recognition of various spaces/perspectives repeatedly and instantly**. Every time a user switches a working space from one display to another (e.g., one display to another that shows various perspectives from multiple virtual avatars or remote robots), the user recognizes the new spatial coordinate as well as the relationship with their own body projections (e.g., avatars or mouse cursors). Here, our research question is: How can user switch between multiple self-bodies in different spaces/perspectives repeatedly and instantly without cognitive load?

Studies have indicated that multitasking body movements can be a highly challenging task because of the limited resources of human awareness and cognitive processes [29, 35, 36]. This phenomenon could be attributed to frequent changes in our internal model (e.g., critical models to generate body interaction strategy including spatial context such as coordinates and body-space relationships, force field, object visual appearance, and so on) when trying to instantly recognize different contexts and objects, rendering the quick selection of an appropriate internal model difficult [15]. Thus, the user needs to select the model every time the spatial context changes in the middle of the observation.

While several studies have focused on overcoming this challenge by showing spatial contexts beforehand as prior learning [5, 18],

it is also known that *presenting pre-attentive information* about visual targets across the entire visual field can still preserve users' awareness toward the targets, allowing the users to process relevant information simultaneously [30, 49].

Given the effect of the pre-attentive visual information on the simultaneous information processing and the attribute of spatial context recognition, we hypothesized that **showing a user's self-body projection to multiple spaces *in unattended areas* could help achieve smoother spatial context recognition**, allowing seamless switching between multiple spaces.

2 OUR APPROACH: SHADOW CLONES INTERFACE

To achieve this, we designed and evaluated a new user interface, called *Shadow Clones*, that projects self-body information in unattended areas for increasing the awareness of body-space relationships and allowing users to seamlessly switch across different visual perspectives in each working space, as shown in Figure 2(B).

To validate our interface, we implemented a system using multiple perspectives, with mouse cursors as the user's body projection. Since the scope of this paper is to investigate performance in the simplest setup for eliminating confounding factors such as muscle fatigue and the visual appearance of the virtual body, we designed our task that uses a mouse cursor based on experimental paradigms in previous research about multiple object tracking [37] and motor control task for investigating internal motor models [20].

In our spatial interface, multiple spaces were arranged in a segmented layout on display. Users can rapidly switch a focused space to interact by looking at the view based on the user's eye gaze (Figure 3(A)). This enables faster switching of the focused space compared to the conventional keyboard switching which involves human latency from the perception to action [27]. To preserve the awareness of the projected user's action in multiple spatial coordinates, the shadow clones interface provides two cursor states, "Body" and "Shadow". The "Body" cursor is displayed in the focused view area the user is gazing at, which indicates that the user is interacting in the space, and the "Shadow" cursors are displayed in dimmed color in the rest of the areas where the user is not gazing at, which works just as visualizations of the user's body movement and the user cannot interact through these shadow cursors.

By displaying clones of the cursor in each space where even the user is not gazing, these can **provide unattended visual information of the user's body movements in the peripheral visual field that helps maintain the sense of body-space coordination**, which facilitates faster switching of views and actions.

In this paper, we implemented and explored the proposed approach through a simple visual reaching task with a performance evaluation in terms of task completion time and success rate. We demonstrated superior performance when compared with the condition of having no shadow clones. We subsequently discuss possible mechanisms of improvement as well as two potential scenarios using the shadow clones approach, including new entertainment content for virtual reality e-sports and multiple robot teleoperation such as in a construction site or a disaster site, without compromising operational performance.

3 RELATED WORK

3.1 Multi-Tasking

Numerous studies have attempted to use multiple virtual reality bodies or multiple robotic bodies simultaneously [13, 31, 32, 42, 43]. However, humans are reportedly unsuitable for multitasking [35, 36], in particular, in a series of psychophysical experiments on the psychological refractory period [35], the processes of perception, response selection, and response execution were investigated [10, 14, 28]. Cognitive operations that require central processing (response selection) can only proceed serially, whereas other operations such as perception and response execution can occur parallelly [36, 41]. This study indicates that a bottleneck exists in cognitive operations required for information integration processing [9, 28, 41, 48] Even prolonged training with dual tasks can reduce multitasking costs[11, 22, 40]. Parallel processing is not actually achieved; rather, individual processes are accelerated and subsequently executed sequentially to achieve resulting parallelism [11, 44].

For multitasking with multiple bodies, perceptual phase and motor execution can be performed simultaneously. However, the cognitive judgment phase can remain a serial central bottleneck [28, 41]. Furthermore, motor execution is physically challenging, when the same body part is used for two exercises. Therefore, the realization of multitasking with multiple bodies to computationally assist in speeding up sequential execution through a user interface is a critical problem that should be resolved.

3.2 Pre-Attentive Processing

While studies have pointed out that there are limitations in performing multitasking [35, 36, 41], it is still possible to perform parallel processing at a pre-attentive level. For instance, in a visual search task, even with a certain amount of visual information presented, a specific visual object can be found immediately by a user [33]. This feature search, also known as “disjunctive” or “efficient” search, is a large-scale visual search process involving unique visual features (such as color, shape, orientation, or size). It is also known that this can still be performed in parallel [45], and users are able to track multiple objects (e.g., up to five objects) in their visual field [1, 37].

Furthermore, this can be achieved in a setup, which is spatially parallel, even during pre-attentive processing [8, 30, 49], suggesting that these visual objects and their movements can be easily detected without user attention.

With these characteristics, many visualization techniques leverages this visual pre-attentive processing to organize information, convey complex information and visual inspection [16, 17, 24, 47].

From this perspective, we take advantage of human automatic and parallel visual processing capabilities by providing unattended information about the user's own body movements to explore the effect and the challenges during multitasking in the aforementioned scenarios.

3.3 Gaze Interface

When assuming multiple robot operations or 3D space operations, multiple viewpoint layouts are used by the user to manipulate objects in spaces with distinct spatial coordinates that are seen from these multiple parallel viewpoints. Using these spatially divided view layouts, users should understand the coordinates of each view space and manipulate objects using aspects that represent the user's body, for example, a cursor. This phenomenon requires switching between multiple cursors on display. For instance, a method has been proposed that uses eye-gaze to select and manipulate a cursor that is close to the target from multiple cursors presented on the screen [3, 38] such as Rake Cursor. Alternatively, several methods have been proposed to speed up the reaching of the desired GUI object by displaying multiple cursors on a screen as Ninja cursors [23]. By extending multi-cursor techniques to virtual reality, Ninja hands facilitate the selection of spatially located 3D objects by arranging multiple virtual hands as a manipulation interface in virtual reality [39]. Another eye-tracking-based interface was proposed for selecting a target window for operation [25], and this method is generally effective because the target window with a keyboard or other interfaces need not be selected.

Based on previous studies, we applied the following two ways to realize the multi-task with multiple bodies to assist in speeding up sequential execution: (1) we provide unattended information about all of the user's multiple bodies simultaneously; (2) we used the gaze to switch each body smoothly.

4 USER STUDY

In this user study, we investigated whether shadow clone can improve task performance by maintaining the sense of body-space coordination of multiple spaces. Our studies were approved by our Institutional Review Board (IRB:22-F-0003).

4.1 Task Design

In this study, we followed the usual reaching task procedure but we applied a successive design with a time limitation to complete the reaching task. We intentionally included this time limit in the reaching task because the proposed approach aims to enable parallel task execution in a fast manner. Figure 3(B) depicts the schematic of the temporal pressure design for the successive reaching task. As shown in Figure 3(C), in this task, a target appeared at every predefined target generation interval time (gen-interval) in a randomly selected different view area; this target appeared only in the view areas in which no other target existed. This task design required the participants to switch between view areas every time a reaching action was performed.

Participants were asked to reach the target before it disappeared in the four viewing areas. Since the objective of this study is to evaluate performance when switching between multiple spaces. Therefore, participants do not necessarily reach the targets in the order of appearance and are allowed to reach the targets in any order. From our preliminary evaluation, we selected 600 ms and 800 ms gen-intervals as two well-balanced levels of difficulty to investigate how temporal pressure affects performance. The time limit of the targets was set as 1000 ms. The rotation of the plane in each of the four spaces was randomly determined. In this experiment, we

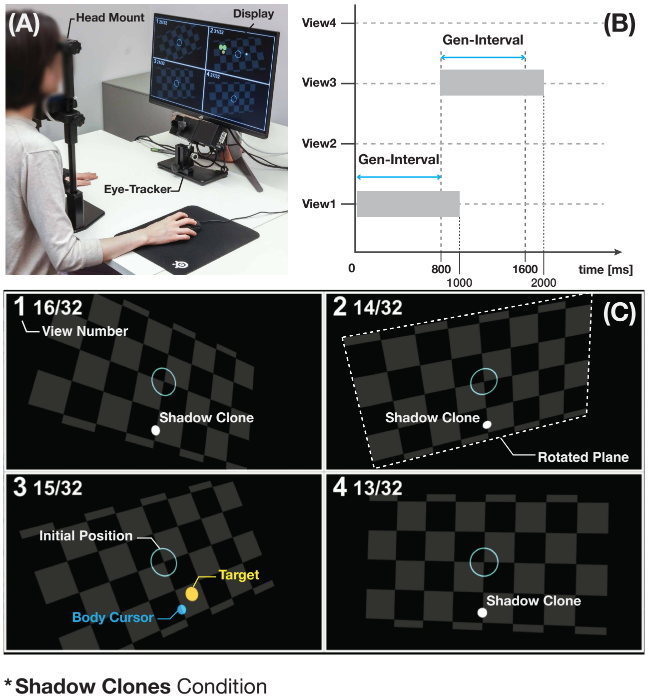

Figure 3: (A) Experimental setup. (B) Target generation sequence at a generating interval time of 800ms. (C) Task setup. Participants need to reach to a dot target within a time limit.

compared the two presentations of the cursor for the reaching task, namely no-shadow, and shadow-clones.

In the **no-shadow** condition, the cursor was only visible with blue color in the view area on which the participant is focusing. In the **shadow-clones** condition, as depicted in Figure 3(C), the focused view area was switched by eye-gazing, which is identical to that in the no-shadow condition, but cursors were also presented in the viewing areas not focused on. The cursor of the current area was colored blue, and the cursors in the other areas were colored white. The local positions of all the cursors were projected and synchronized with the mouse, regardless of whether the view area was being focused on or not.

4.2 Procedure

First, participants were instructed about the gen-interval and the time limitation in our reaching task, as well as the switching method using the eye-gaze. The calibration was performed by the operator before the experiment. In both conditions, participants practiced 12 trail (3 times on the plane in each space) prior to the actual experiment. After the experimenter explained the procedure to the participants, they performed the experiment, which consisted of four sessions, two in the gen-interval condition and two cursor conditions. There were breaks between each session. We checked whether eye gaze was accurately detected before the session. The number of trials for each view area was 32, and 512 (32 trials* 4 views * 4 conditions) trials were conducted for the two gen-interval conditions and two cursor conditions for a total of four conditions. The experiment was conducted with eight counterbalances.

4.3 Participants

We recruited 14 participants (nine self-identified as female, seven as male) from a local institute and from outside using a mailing list. Their ages ranged from 18 to 37 years (mean = 25, SD = 4.2), and all had normal or corrected normal vision. Participants were compensated with 3000JPY for their time. Data of two participants whose visual acuity was less than 0.1 were excluded because the participants could not clearly view the target or cursor.

4.4 Results

We collected the participants' performances, including the number of successes in reaching the target and the task completion time in the reaching tasks for each session.

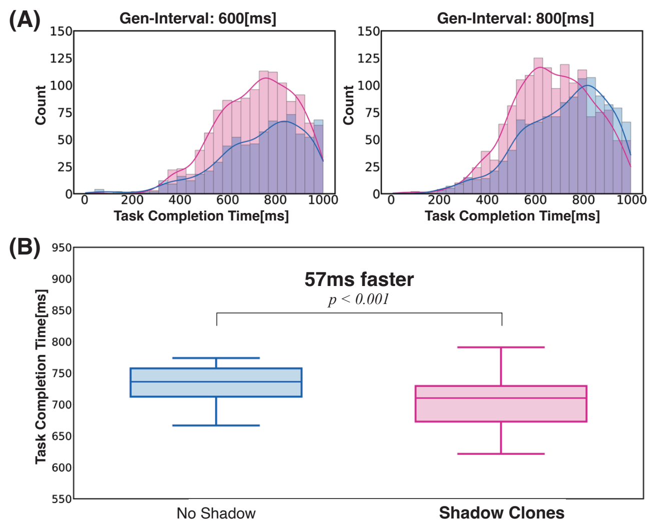

Figure 4: Distribution of the task completion time in (A) 600 ms and 800 ms gen-interval. (B) Task completion time in the no-shadow and the shadow-clones conditions.

4.4.1 Task Completion Time. The general distribution of the task completion time, that is, the duration from when a target appeared to when it was reached, revealed faster trends for the shadow-clones condition in the user study (Figure 4(A) and (B)). To validate the differences, we performed a statistical analysis of the median of the task completion time between the no-shadow and the shadow-clones conditions using a two-way analysis of variance (ANOVA). Because this study focused on the task completion time across the view, including view switching, we excluded the data from trials of reaching without view switching, and also excluded the data in failure of reaching trials which we analyze later on.

The results revealed a significant difference in the cursor-view conditions; the no-shadow (600 ms gen-interval: Mdn = 770.833, SD. = 181.754, 800 ms gen-interval: Mdn = 750.0, SD. = 174.888) and the shadow-clones (600 ms gen-interval: Mdn = 737.5, SD. = 161.311, 800 ms gen-interval: Mdn = 675.0, SD. = 168.898), in total

F(1,13) = 18.2, $p < 0.001$, $\eta_p^2 = 0.584$. No statistical difference was observed between the gen-interval conditions, and no interaction was observed between the cursor-view and the gen-interval conditions. On the basis of these results, we calculated the gain for the shadow clones; the gain was defined as the difference in the medians of the task completion time, that is, how much faster the participants got in the shadow-clones condition than that in the no-shadow condition. We found that the task completion time in the shadow-clones condition was 57 ms faster (Figure 4(C)). Given the higher task success rate, we can consider 57 ms to be a significant gain for the shadow-clones interface compared with using only a single cursor for the task completion process.

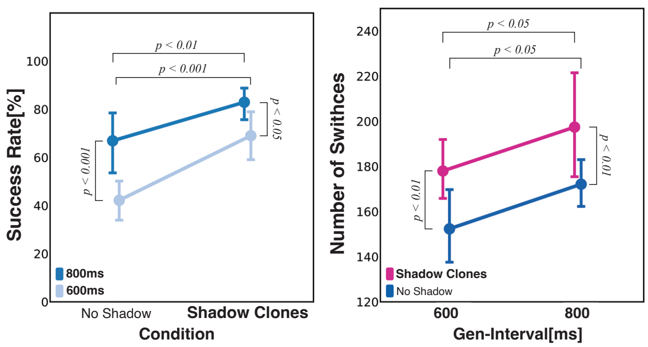

Figure 5: (LEFT) Success rate: shadow-clones condition outperformed the no-shadow condition. (RIGHT) Number of the view switching in two gen-intervals: Significant differences between no-shadow and shadow-clones were observed.

4.4.2 Success Rate. Figure 5(A) displays the success rates. To investigate the differences in the two cursor conditions (no shadow and shadow clones) and the two gen-intervals for the targets (600 and 800 ms), a two-way repeated ANOVA analysis was conducted on the success rate of each switching condition and two gen-intervals. The result showed that a significant difference was observed between the cursor conditions (F(1, 13) = 37.4, $p < 0.001$, $\eta_p^2 = 0.742$). A post-hoc test on the cursor condition based on the Holm-Sidak test revealed that the success rates for the 600 ms gen-interval differed significantly between cursor conditions (no-shadow: Mdn = 41.02, SD. = 16.26, shadow-clones: Mdn = 73.44, SD. = 19.41, $p < 0.001$), and a significant difference also existed for the 800 ms gen-interval (no-shadow: Mdn = 74.22, SD. = 24.8, shadow-clones: Mdn = 87.11, SD. = 13.35, $p < 0.01$). A post-hoc test on the gen-interval based on the Holm-Sidak test revealed that the success rates for the no-shadow condition differed significantly between the gen-intervals ($p < 0.001$), and a significant difference ($p < 0.05$) was also observed in the success rates for the shadow-clone condition.

4.4.3 Perspective Switching. Significant differences (F(1,13) = 12.7, $p < 0.01$) were observed in the number of view switching in the gen-interval conditions between no-shadow (600 ms gen-interval: Mdn = 152.36, SD. = 31.27, 800 ms gen-interval: Mdn = 166, SD. = 21.62) and shadow-clones (600 ms gen-interval: Mdn = 176.5, SD. = 26.23, 800 ms gen-interval: Mdn = 184.5, SD. = 47.3) condition. (Figure 5(B)). The no-shadow and shadow-clones conditions were compared at each gen-interval, and significant differences were also observed. This also supports how the shadow clones facilitate faster view switching in our successive reaching task. These results clearly demonstrated that the participant performance was better with shadow clones than with no shadows.

5 DISCUSSION

The proposed interface with gaze-switching and shadow clones leads to superior performance in the successive reaching tasks with multi-spatial coordinate views. The superiority of the proposed shadow clones was observed in two different time pressure difficulties. Here, we discussed the cause of the improved performance of the participants in serial actions with multiple viewing areas.

To complete the designed reaching task, participants required several processing steps: 1) detecting newly appearing targets, 2) orienting their eye gaze to see the targets, 3) switching the view areas to reach the next target, and 4) recalibrating their mouse movement in the new current view area and subsequently reaching the target. As we mentioned, eye-gaze-based view-switching can be performed without the cognitively demanding processes. Eye-gaze-based view switching enables fast switching and interaction across multiple spaces by leveraging visual pre-attentive processes [33, 49] that humans inherently possess and can be processed in parallel [30, 37, 41]. Therefore, process (3) can be automatically accomplished by leveraging (1) and (2) behaviors by eye-gaze switching.

5.1 Cause of Superior User Performance

In the user study, the superior performance can be attributed to the shadow clone interface, in which the cursor positions were visualized across all view areas at multiple coordinates. Although several factors influence the task completion time, an additional analysis was performed to investigate the gain in the completion time.

To determine the cause of users becoming faster with shadow clones, we first considered how the system delay from eye-to-display can explain the time. We recorded the actual eye movement and the display output using a high-speed camera [1] at 1000 Hz. From the recorded video, we measured the eye-display latency from the eye movement to the display update and found that it was approximately 20 ms. Furthermore, the recorded eye-tracking stream data indicated that the saccade movement between the view areas was approximately 60 ms. This result is consistent with those reported by previous literature [2, 7] on the saccade movement. With a 20-degree movement, by considering the distance on the screen, the expected saccade duration was calculated to be 60-80 ms.

Because the view switching occurs when the eye gaze goes beyond the center line, it occurred 30 ms after the saccade started. Considering the saccade duration, the view area that is currently focused on may have switched before the saccade was completed. This result is consistent with the subjective reports from some participants; the participants felt the cursor was originally blue even before the gaze switch.

With saccadic suppression, which is a well-known phenomenon in which human vision is suppressed during saccade movement,

[1]XIMEA MQ013CG-ON

participants may not notice any system delay due to saccade suppression depending on the saccade distance. This phenomenon suggests that the gain of 57 ms in the task completion time was not attributed to the system delay alone but also to the shadow clones, in which all four cursor clones were displayed in multiple viewing areas.

5.2 User Interface Leveraging Implicit Process

The shadow clone interface suggests the existence of an implicit cognitive process for performing tasks across multiple spaces by displaying cursors that projects the self-body onto the view areas that the user is not gazing at. Figure 5(B) displays that the presence of a cursor that projects self-motion in each space implicitly allows the user to grasp the relationship between self-motion and position in that space and to maintain the physicality of each space.

In contrast to previous approaches such as Ninja Hands [39] and Ninja cursors [23] that allow user to preserve the sense of body movement to multiple operating objects in the single space coordination, our shadow clones allow users to preserve the sense of body movement to multiple operating objects in multiple corresponding body-space coordination. In addition, from the findings of Rake Cursor [3] that allows users to rapidly switch the interfaces to control with eye-gaze, we applied eye-gaze based switching across multiple spaces to enable users to experience multiple interactions simultaneously.

By integrating those approaches, the implicit processes of preserving multiple visual coordinates could potentially enable the immediate spatial manipulation of the cursor across multiple body-space coordinations.

6 POTENTIAL APPLICATIONS AND FUTURE WORK

We proposed two potential applications to demonstrate the use of shadow clone interface and highlight their potential for the enhancement of the capabilities of virtual reality and robotic systems with multiple bodies in multiple spaces.

6.1 Interaction of Multiple Bodies using the VR Body

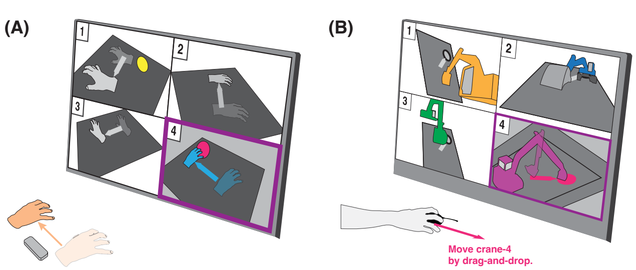

Figure 6: Potential applications: Shadow clones can be applied for controlling (A) multiple virtual hands or (B) multiple construction machines, enabling a single user to control distributed robots in a remote place.

With the virtual reality body, users have the potential to control multiple bodies, not only just one body, which will create novel interactions for entertainment and learning [31, 46]. Since the user's perspective can be arbitrarily controlled in three-dimensional games [19], a sense of multiple spatial coordination is required to control virtual bodies in multiple game worlds.

Our proposed shadow clones allow us to extend those multiple virtual body scenarios to multiple visual coordinates. For instance, the user will be able to control the virtual hands of multiple characters in different spaces and even play multiple different games in parallel(Figure 6(A)).

6.2 Multiple Action using Multiple Robots

The use of multiple robots can increase in various applications such as construction and disaster response [26]. In these situations, a single operator can control multiple robots from multiple perspectives, which can reduce manpower resources and improve efficiency. This method can enhance the capabilities and versatility of robots and allow for more complex tasks to be performed simultaneously. Previous research proposed several approaches to switch the operating robot arm from multiple robots such as Parallel Ping-Pong project [43] with the system-triggered switch or users physiology sensing-based switch [32]. Besides those previous approaches only expected to initially visual coordinate, i.e., the same camera-arm coordination, our shadow clones will extend those use cases to more arbitrary situations, including multiple perspectives (Figure 6(B)). The shadow clones will provide a seamless and intuitive way to control multiple robots simultaneously. This can improve the efficiency and effectiveness of construction and other tasks involving multiple robots.

6.3 Limitations and Future Work

For applying Shadow Clones to physical multiple bodies, such as robotic arms, it will be an important challenge to consider physical interference caused by unintended motion when we move all bodies simultaneously. To overcome this limitation, we expect that methods of robotic control using augmented reality [4, 6, 21, 34], such as visualizing user inputs as a virtual representation with real images, can also be integrated into our Shadow Clones approach to mediate the inconsistency between the virtual and physical situation. In addition, when the user controls a robot or machine, instead of constantly controlling the manipulator (e.g., robot's hand, the crane of construction machine), it is controlled by drag-and-drop while concentrating on the controlling target to prevent unintended motion (Figure 6(B)).

Furthermore, our quantitative findings from our user study bring us to motivate to evaluate the user's subjective experience. However, the number of participants in the user study was small, and the task design was limited to a simple reaching task. In the future, we will investigate the performance of our approach when more bodies are present through many participants and other task design, improve the robustness of the sense of body-space coordination as well as the cognitive load for multiple bodies, and aim to elucidate the mechanism of parallel interaction.

We used a cursor as the simplest representation of the user's body in our experiment to investigate visual-movement coordination from multiple visual perspectives. Further research questions include how body representations affect the adaptability of multiple bodies and multiple visual perspectives as well as the sense of body ownership. Through this research, we aim to gain insights into the design of multi-body embodied applications in the future.

7 CONCLUSION

We designed a Shadow Clones interface that can display multiple cursors across all view areas in various body-space coordinations; in this interface, the cursor can be activated when the respective area is focused on. This interface will allow users to maintain multiple senses of body-space coordination from multiple visual perspectives by constantly presenting body projections of self-movement. Our results of the user study revealed that the proposed interface improved performance and facilitated faster task completion in the successive reaching task with multi-spatial coordinate view areas. Our study findings provide a foundation for designing interfaces that will enable users to perform multiple tasks not only with a screen-based interface but also with multiple bodies in multiple worlds.

ACKNOWLEDGMENTS

This work was supported by JST Moonshot R&D Program (Grant Number JPMJMS2013).

REFERENCES

[1] Reem Alzahabi and Matthew S Cain. 2021. Ensemble perception during multiple-object tracking. *Atten. Percept. Psychophys.* 83, 3 (April 2021), 1263–1274.

[2] A T Bahill, A Brockenbrough, and B T Troost. 1981. Variability and development of a normative data base for saccadic eye movements. *Invest. Ophthalmol. Vis. Sci.* 21, 1 Pt 1 (July 1981), 116–125.

[3] 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)*. Association for Computing Machinery, New York, NY, USA, 1415–1418.

[4] Yuanzhi Cao, Tianyi Wang, Xun Qian, Pawan S Rao, Manav Wadhawan, Ke Huo, and Karthik Ramani. 2019. GhostAR: A Time-space Editor for Embodied Authoring of Human-Robot Collaborative Task with Augmented Reality. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 521–534.

[5] Li-Hung Chang, Kazuhisa Shibata, George J Andersen, Yuka Sasaki, and Takeo Watanabe. 2014. Age-related declines of stability in visual perceptual learning. *Curr. Biol.* 24, 24 (Dec. 2014), 2926–2929.

[6] Linfeng Chen, Kazuki Takashima, Kazuyuki Fujita, and Yoshifumi Kitamura. 2021. PinpointFly: An Egocentric Position-control Drone Interface using Mobile AR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21, Article 150)*. Association for Computing Machinery, New York, NY, USA, 1–13.

[7] H Collewijn, C J Erkelens, and R M Steinman. 1988. Binocular co-ordination of human horizontal saccadic eye movements. *J. Physiol.* 404 (Oct. 1988), 157–182.

[8] Barbara Anne Dosher, Songmei Han, and Zhong-Lin Lu. 2010. Information-limited parallel processing in difficult heterogeneous covert visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 5 (Oct. 2010), 1128–1144.

[9] John Duncan. 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* 14, 4 (April 2010), 172–179.

[10] Paul E Dux, Jason Ivanoff, Christopher L Asplund, and René Marois. 2006. Isolation of a central bottleneck of information processing with time-resolved FMRI. *Neuron* 52, 6 (Dec. 2006), 1109–1120.

[11] Paul E Dux, Michael N Tombu, Stephenie Harrison, Baxter P Rogers, Frank Tong, and René Marois. 2009. Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron* 63, 1 (July 2009), 127–138.

[12] Dylan F Glas, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2012. Teleoperation of Multiple Social Robots. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 42, 3 (May 2012), 530–544.

[13] Arvid Guterstam, Dennis E O Larsson, Joanna Szczotka, and H Henrik Ehrsson. 2020. Duplication of the bodily self: a perceptual illusion of dual full-body ownership and dual self-location. *R Soc Open Sci* 7, 12 (Dec. 2020), 201911.

[14] Suk Won Han and René Marois. 2013. The source of dual-task limitations: serial or parallel processing of multiple response selections? *Atten. Percept. Psychophys.* 75, 7 (Oct. 2013), 1395–1405.

[15] James B Heald, Máté Lengyel, and Daniel M Wolpert. 2021. Contextual inference underlies the learning of sensorimotor repertoires. *Nature* 600, 7889 (Dec. 2021), 489–493.

[16] Christopher G Healey, Kellogg S Booth, and James T Enns. 1996. High-speed visual estimation using preattentive processing. *ACM Trans. Comput. Hum. Interact.* 3, 2 (June 1996), 107–135.

[17] Christopher G Healey and James T Enns. 2012. Attention and visual memory in visualization and computer graphics. *IEEE Trans. Vis. Comput. Graph.* 18, 7 (July 2012), 1170–1188.

[18] Yoko Higuchi, Yoshiyuki Ueda, Kazuhisa Shibata, and Jun Saiki. 2019. Spatial variability induces generalization in contextual cueing. *J. Exp. Psychol. Learn. Mem. Cogn.* (Nov. 2019).

[19] Matthias Hoppe, Andrea Baumann, Patrick Chofor Tamunjoh, Tonja-Katrin Machulla, Paweł W Woźniak, Albrecht Schmidt, and Robin Welsch. 2022. There Is No First- or Third-Person View in Virtual Reality: Understanding the Perspective Continuum. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22, Article 360)*. Association for Computing Machinery, New York, NY, USA, 1–13.

[20] Hiroshi Imamizu and Mitsuo Kawato. 2009. Brain mechanisms for predictive control by switching internal models: implications for higher-order cognitive functions. *Psychol. Res.* 73, 4 (July 2009), 527–544.

[21] Shunichi Kasahara, Ryuma Niiyama, Valentin Heun, and Hiroshi Ishii. 2013. exTouch: spatially-aware embodied manipulation of actuated objects mediated by augmented reality. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction* (Barcelona, Spain) *(TEI '13)*. Association for Computing Machinery, New York, NY, USA, 223–228.

[22] A M Clare Kelly and Hugh Garavan. 2005. Human functional neuroimaging of brain changes associated with practice. *Cereb. Cortex* 15, 8 (Aug. 2005), 1089–1102.

[23] 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)*. Association for Computing Machinery, New York, NY, USA, 949–958.

[24] Andrey Krekhov and Jens Krueger. 2021. Deadeye: A Novel Preattentive Visualization Technique Based on Dichoptic Presentation. (Jan. 2021). arXiv:2101.07048 [cs.HC]

[25] Manu Kumar, Andreas Paepcke, and Terry Winograd. 2007. EyeExpos'e: Switching applications with your eyes.

[26] Seungyeol Lee and Jeon-Il Moon. 2017. Management of a Single-User Multi-Robot Teleoperated System for Maintenance in Offshore Plants. In *Multi-agent Systems*. InTech.

[27] Bingxin Li, Xiangqian Li, Gijsbert Stoet, and Martin Lages. 2019. Exploring individual differences in task switching. *Acta Psychol.* 193 (Feb. 2019), 80–95.

[28] René Marois and Jason Ivanoff. 2005. Capacity limits of information processing in the brain. *Trends Cogn. Sci.* 9, 6 (June 2005), 296–305.

[29] Antonella Maselli and Mel Slater. 2014. Sliding perspectives: dissociating ownership from self-location during full body illusions in virtual reality. *Front. Hum. Neurosci.* 8 (Sept. 2014), 693.

[30] B McElree and M Carrasco. 1999. The temporal dynamics of visual search: evidence for parallel processing in feature and conjunction searches. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 6 (Dec. 1999), 1517–1539.

[31] Reiji Miura, Shunichi Kasahara, Michiteru Kitazaki, Adrien Verhulst, Masahiko Inami, and Maki Sugimoto. 2022. MultiSoma: Motor and Gaze Analysis on Distributed Embodiment With Synchronized Behavior and Perception. *Frontiers in Computer Science* 4 (2022).

[32] Yukiya Nakanishi, Masaaki Fukuoka, Shunichi Kasahara, and Maki Sugimoto. 2022. Synchronous and Asynchronous Manipulation Switching of Multiple Robotic Embodiment Using EMG and Eye Gaze. In *Augmented Humans 2022* (Kashiwa, Chiba, Japan) *(AHs 2022)*. Association for Computing Machinery, New York, NY, USA, 94–103.

[33] Ulric Neisser and Robert Becklen. 1975. Selective looking: Attending to visually specified events. *Cogn. Psychol.* 7, 4 (Oct. 1975), 480–494.

[34] Mikhail Ostanin, Stanislav Mikhel, Alexey Evlampiev, Valeria Skvortsova, and Alexandr Klimchik. 2020. Human-robot interaction for robotic manipulator programming in Mixed Reality. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 2805–2811.

[35] H Pashler. 1994. Dual-task interference in simple tasks: data and theory. *Psychol. Bull.* 116, 2 (Sept. 1994), 220–244.

[36] H Pashler. 1994. Graded capacity-sharing in dual-task interference? *J. Exp. Psychol. Hum. Percept. Perform.* 20, 2 (April 1994), 330–342.
[37] Z W Pylyshyn and R W Storm. 1988. Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat. Vis.* 3, 3 (1988), 179–197.
[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)*. Association for Computing Machinery, New York, NY, USA, 1411–1414.
[39] 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*. Association for Computing Machinery, New York, NY, USA, 1–14.
[40] E H Schumacher, T L Seymour, J M Glass, D E Fencsik, E J Lauber, D E Kieras, and D E Meyer. 2001. Virtually perfect time sharing in dual-task performance: uncorking the central cognitive bottleneck. *Psychol. Sci.* 12, 2 (March 2001), 101–108.
[41] Mariano Sigman and Stanislas Dehaene. 2008. Brain mechanisms of serial and parallel processing during dual-task performance. *J. Neurosci.* 28, 30 (July 2008), 7585–7598.
[42] Kazuma Takada, Midori Kawaguchi, Yukiya Nakanishi, Akira Uehara, Mark Armstrong, Adrien Verhulst, Kouta Minamizawa, and Shunichi Kasahara. 2021. Parallel Ping-Pong: Demonstrating Parallel Interaction through Multiple Bodies by a Single User. In *SIGGRAPH Asia 2021 Emerging Technologies* (Tokyo, Japan) *(SA '21 Emerging Technologies, Article 12)*. Association for Computing Machinery, New York, NY, USA, 1–2.
[43] Kazuma Takada, Midori Kawaguchi, Akira Uehara, Yukiya Nakanishi, Mark Armstrong, Adrien Verhulst, Kouta Minamizawa, and Shunichi Kasahara. 2022. Parallel Ping-Pong: Exploring Parallel Embodiment through Multiple Bodies by a Single User. In *Augmented Humans 2022* (Kashiwa, Chiba, Japan) *(AHs 2022)*. Association for Computing Machinery, New York, NY, USA, 121–130.
[44] James T Townsend. 1990. Serial vs. Parallel Processing: Sometimes They Look like Tweedledum and Tweedledee but they can (and Should) be Distinguished. *Psychol. Sci.* 1, 1 (Jan. 1990), 46–54.
[45] A M Treisman and G Gelade. 1980. A feature-integration theory of attention. *Cogn. Psychol.* 12, 1 (Jan. 1980), 97–136.
[46] Adrien Verhulst, Yasuko Namikawa, and Shunlchl Kasahara. 2022. Parallel Adaptation: Switching between Two Virtual Bodies with Different Perspectives Enables Dual Motor Adaptation. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 169–177.
[47] C Ware. 2004. *Information Visualization: Perception for Design: Second Edition*. unknown.
[48] Kei Watanabe and Shintaro Funahashi. 2014. Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nat. Neurosci.* 17, 4 (April 2014), 601–611.
[49] Jeremy M Wolfe. 2014. Approaches to Visual Search. In *The Oxford Handbook of Attention*, Anna C. (Kia) Nobre and Sabine Kastner (Eds.). Oxford University Press.