Designing with Hands: Hand-Tracking Interaction for Urban and Landscape Design in Virtual Reality

Mengdie Wang[1]

[1]Nanyang Technological University Singapore · MENGDIE001@e.ntu.edu.sg

Abstract: Hand tracking enables more direct, embodied interaction in virtual reality (VR), yet its practical value for creating urban and landscape design interventions remains under-examined. This paper presents a VR simulator that enables users to explore a site and propose spatial interventions through hand-tracking interaction, transforming design into an embodied act of spatial exploration and making. A pilot user study (N=5) examined the simulator's usability and user experience. Results suggest that hand tracking supports the full interaction and provides practical advantages during immersive design activity. This paper contributes a hand-tracking-enabled VR interaction approach and derives design implications for accessible urban and landscape design with lay participants.

Keywords: Virtual Reality, hand tracking, simulator, participatory design

1 Introduction

Virtual reality (VR) has become an increasingly important medium for urban and landscape design, offering immersive spatial exploration and interactive visualization that can extend beyond professional studios toward broader public engagement (PORTMAN et al. 2015). In participatory design contexts, however, the value of VR depends not only on what is visualized, but on who can meaningfully act within the environment. The ways people interact in virtual space shape how they understand spatial conditions, express preferences, and translate everyday needs into design moves. Conventional VR systems typically rely on handheld controllers mapped to button inputs, which convert spatial intent into abstract operations. This separation can introduce a cognitive and practical gap between bodily movement and spatial action, raising the interaction threshold for non-experts and potentially creating a barrier for participants without prior controller familiarity.

Hand-tracking technology presents an alternative paradigm that aligns more closely with participatory goals. By recognizing and interpreting natural hand gestures, it reduces reliance on intermediary devices and supports a more direct form of physical authorship in digital space (BUCKINGHAM 2021). For lay participants, this can lower the entry barrier by allowing interaction to build on familiar embodied skills: pointing, reaching, placing, and gesturing, rather than on learned button mappings. As a result, hand tracking may help participants engage more confidently, sustain attention on spatial decision-making rather than interface translation, and contribute design interventions that better reflect lived experience and situated knowledge.

This paper presents a VR simulator and hand-tracking interaction framework for urban and landscape design interventions. The framework treats navigation as a foundational act of spatial perception and supports four interaction methods: gesture-based vegetation planting, urban feature placement and manipulation, creative placeholders with semantic annotation, and sphere-triggered surface editing. The system operationalizes these methods through a design toolkit and contextual editing mechanisms, and it is evaluated in a formative qualitative pilot

Journal of Digital Landscape Architecture, 11-2026, pp. 306-313. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-770-0, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537770029.

that compares hand tracking and controller input. Observation and post-session interviews were used to clarify how hand tracking supports design actions in VR and to examine its participatory and educational potential. This work contributes a validated simulator prototype and empirical evidence that characterizes modality trade-offs, informing design implications for future participatory design tools.

2 From Immersive Visualization to Embodied Design in VR

Immersive visualization has long been used in architecture and urban design to communicate proposals, test spatial scenarios, and support stakeholder understanding (PORTMAN et al. 2015). In participatory planning, however, many tools still privilege consultation over co-creation, where non-experts respond to options rather than shape them. SANDERS (2002) articulated this need of to shift from "designing for users" to "designing with users," moving beyond user-centred approaches toward participatory experiences where stakeholders become active co-creators. VR can shift this balance by enabling participants to not only see alternatives but also compose them by placing elements, adjusting configurations, and iterating proposals in situ (CHOWDHURY & SCHNABEL 2019, VAN LEEUWEN et al. 2018), bringing participatory input closer to design activity rather than feedback alone.

This shift matters because architectural and urban understanding is fundamentally experiential. Place is grasped through movement, changing viewpoints, and bodily presence (MALININ 2016), and design judgment often develops through iterative reflection while acting within a situation (DOURISH 2001, SCHÖN 2017). For VR-based participatory design, this implies that interaction should support multi-perspectival navigation and direct interaction and manipulation with the virtual environment. From this perspective, hand tracking may lower the participation threshold for non-experts by aligning interaction with familiar bodily skills (BUCKINGHAM 2021), helping participants concentrate on spatial intent and design needs.

3 VR Simulator Design and Interaction Framework

3.1 System Overview

A VR simulator was developed in Unity using Meta's All-in-One SDK for hand-tracking implementation and was tested on a Meta Quest 3 headset. The simulator reconstructs the residential district centred on the upcoming transit station at Nanyang Technological University, including surrounding student plazas and green spaces. The intent is not hyper-realistic replication, but a responsive and coherent spatial setting that supports immersion and meaningful creation and editing actions. The simulator features a three-level structure and includes a central design toolkit for design interactions.

3.2 Navigation and Locomotion

The simulator provides three hand-gesture-based locomotion modes: teleportation, continuous sliding, and snap turning. Teleportation uses a pinch gesture: users stretch the index finger and thumb to indicate a destination and pinch to move instantly. Continuous movement

is controlled through a thumb-up gesture to slide forward and a thumb-down to move backward. Snap turning allows users to rotate their viewpoint using a pinch gesture on the directional turning arrow, triggered by extending the thumb and index finger. A hand menu is attached to the user's left hand to access the design toolkit and system control (Figure 1).

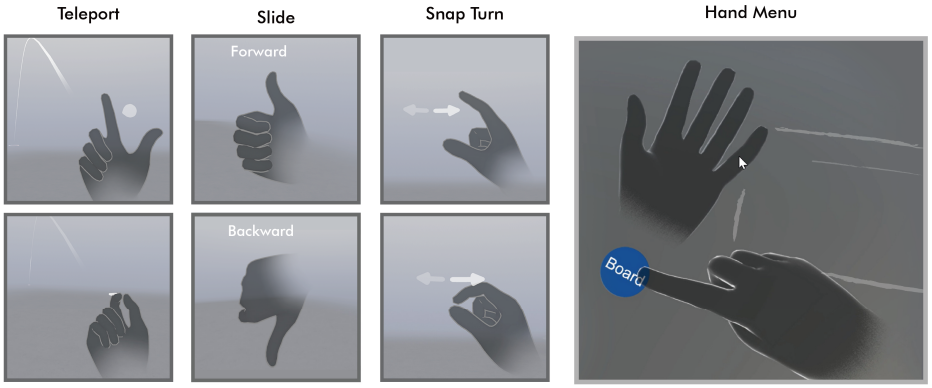

Fig. 1: Hand gestures for Navigation (Teleport, Slide and Snap Turn). Hand Menu is attached to the left hand and toggled through the wrist button.

3.3 The Design Toolkit

The simulator incorporates a layered design toolkit that enables users to make spatial interventions and modify the virtual environment. The toolkit consists of two integrated components: a hand-mounted menu containing eight categories of urban assets, and scene-embedded triggers covering two additional asset categories. Each category includes associated assets and editable properties, including style variants, transforms (position, rotation, scale), materials, and open-end annotations. Figure 2 shows the hand-mounted design toolkit.

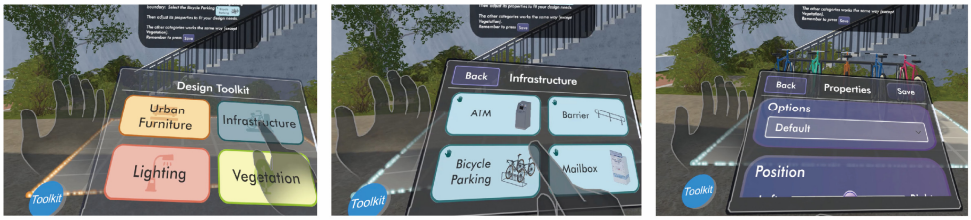

Fig. 2: Hand-mounted toolkit UI: category (left), asset (middle), and property (right)

The hand-menu assets include categories such as Vegetation (planted within designated area), Urban Furniture, Infrastructure, Installation Art, Playground, info, Lighting and Customized, which can be spawned and repositioned directly within the scene. The scene-embedded triggers support Pavement and Architectural Façade (allowing users to annotate indoor space) modification. These assets operate at a larger spatial scale, focusing on material transformation and spatial function changes.

3.4 Design Interaction Framework

The simulator supports full hand-tracking interaction through four complementary paradigms with the design toolkit and representative in-VR screenshots are shown in Figure 3.

Fig. 3: 1) Gesture-based Vegetation Planting (top left), 2) Urban Feature Direct Placement (top right, The UI interaction is shown in Figure 2, this image illustrates direct hand-grab placement), 3) Creative placeholder with semantic annotation (bottom left), 4) Sphere-triggered surface editing (bottom right)

1) Gesture-based Vegetation Planting – Users use a pinch gesture (thumb and index finger) to select vegetation planting areas on the grass field; pinching again deselects the area. Once an area is selected, users choose a vegetation asset from the hand menu toolkit, and plant it within the designated zone.

2) Urban Feature Direct Placement – Users select assets via buttons on the hand menu; the selected asset is instantiated five meters in front of the camera, and the hand menu opens the property interface to adjust its parameters. Instantiated assets can also be manipulated directly via hand interaction: users grab an object with one hand while navigating with the other to reposition and align it within the environment. During grabbing, assets are automatically scaled to a hand-held proportion to reduce visual occlusion.

3) Creative placeholder with semantic annotation – Users can press the Customized button to place a cube as a placeholder to propose a specific asset not available in the toolkit. Besides transformation and material adjustments, they can annotate it via a virtual keyboard to label intended functions or design ideas directly in the virtual scene.

4) Sphere-triggered surface editing – Users touch floating spheres placed in the environment to open contextual UI panels for surface customization, such as changing pavement materials. This interaction encourages focused attention on ground texture, continuity, and material composition in public-space design. The same approach is applied to architectural

façade modification, which additionally includes an annotation feature that allows users to type ideas or proposed functions for interior spaces directly within the virtual scene.

All interactions are also mapped to controllers for equivalent functionality.

3.5 Levels and Procedure

Level 1 (Explore) supports spatial familiarization through navigation. Participants traverse the virtual site to four designated checkpoints to build spatial understanding.

Level 2 (Guided challenges) provides four framework-aligned tasks as a tutorial. It scaffolds use of the design toolkit, with each interaction method introduced as one guided challenge.

Level 3 (Sandbox design) is an open-ended phase that integrates Levels 1 and 2 into a single design loop. Participants freely navigate to assess spatial conditions and implement interventions with the toolkit.

4 Pilot Study Design and Analysis

A formative qualitative pilot was conducted with five participants from the school community (P1-P5) to assess feasibility and inform refinements to the simulator and its interaction design. Participants varied in prior experience: P1, P2 had both VR and spatial design experience; P3 had VR experience only; and P4, P5 had neither. The session followed a three-level workflow. In Level 3, participants first continued designing with the hand tracking and were then provided with the controller during the middle-to-late stage of the session to try the same interactions and make an immediate comparison.

Duration was recorded for each level. Qualitative data were collected through in-session observation and a post-session interview. Observation focused on learnability and interaction breakdowns as participants navigated, used the hand menu, selected and manipulated assets, and performed contextual editing. The interview elicited comparative reflections on naturalness, ease of use, enjoyment, and frustrations for hand tracking versus controller input, grounded in the interaction paradigms. Participants were also asked what they liked most about hand tracking in the simulator, what they disliked or struggled with most, and to describe one concrete moment from the session that illustrated their view.

Observation notes and interview responses were consolidated after each session and analysed using a rapid qualitative approach. Recurring issues and positive affordances were grouped into themes and organized by interaction component (navigation and the four interaction paradigms) and by input modality (hand tracking vs. controller) to capture context-specific trade-offs. Themes were then synthesized across participants to identify consistent patterns and illustrative moments, and findings are reported as formative insights that informed iterative refinements and motivated design implications.

5 Results

For the five pilot participants, Level 1 duration had a median of 4.95 min (range 3.32-6.77), and Level 2 duration had a median of 9.03 min (range 5.87-14.27). Level 3 duration depends

on participants' performance, with participants typically producing around 8-10 design interventions on average. In this pilot, Level 3 lasted a median of 17.53 min (range 12.98-27.20). All participants completed the full three-level session within approximately 22.17-46.42 min.

In-session observation showed that participants generally became more fluent using hand gesture after a short familiarization period. The most common breakdown under hand tracking was unintended UI activation, where participants mis-touched or mis-selected interface elements when interacting with menus and panels. Gesture recognition was also sensitive to how standard the gesture was performed: snap-turn gestures were sometimes not detected reliably, leading several participants to turn physically instead. Another observed issue was that some participants held a gesture posture even when not intending to trigger an action; when switching to a different gesture, the previous gesture could remain active and interfere with recognition, making mid-task transitions feel effortful. In moments of uncertainty, participants occasionally confused gestures and required reminders, particularly when transitioning between commands. This also happens when participants switched to the controller during Level 3, those with prior VR experience adapted quickly and regained stable interaction with minimal guidance. In contrast, the VR-novice participants required additional time to recall controller mappings and showed more hesitation during early controller use.

Interview responses aligned with the observed patterns and indicated modality-dependent trade-offs across navigation and the four interaction paradigms. Participants with prior VR experience (P1, P3) found the controller faster and more precise, and P1 described it as natural and easy once familiar mappings were learned, particularly for stable selection. For navigation, participants generally reported that the controller was easier, faster, and more precise for moving through the site. Across paradigms that relied on UI elements, participant P5 described hand-based UI interaction as "like a touchscreen," and all participants described hand tracking as more fun and intuitive, especially for directly grabbing and manipulating objects.

At the same time, differences emerged by paradigm. For vegetation area selection (Paradigm 1), all participants indicated that the controller was easier because button press was faster and more reliable than performing a pinch gesture. For the hand-menu-dominated interaction (Paradigm 2), hand tracking was preferred because direct finger-based interaction felt more immediate. For Paradigm 3, participants reported that hand input felt natural for typing but was constrained by stability demands: the hand-mounted keyboard was often described as difficult because the hands could not always be held steady. Similarly, surface editing with annotation in Paradigm 4 also involved text input, and participants reported that hand tracking was easy to use when the keyboard was stable and clearly anchored in the scene.

6 Discussion and Conclusion

The formative pilot indicates that the simulator is feasible for the targeted urban and landscape design tasks. All participants completed the three-level workflow within a single session, and hand-tracking supported the intended operations for spatial navigation, toolkit-based design interventions and contextual editing. Observed issues were primarily matters of tracking accuracy and sensitivity rather than failures that prevented participants from completing design tasks. Modality preference varied with prior VR experience: experienced participants tended to prefer controllers, while VR-novice participants engaged more smoothly

with hand tracking. This suggests that familiarity with controller mappings and confirmation logic influenced perceived ease of interaction across input modalities in this pilot. Even though controller input was only introduced later in Level 3, these VR-novice participants still judged hand tracking as sufficient for completing the design operations, implying that the controller's potential advantage was not decisive in their case.

The simulator underscores the importance of the virtual environment as a medium for urban and landscape design. Experiencing the site at human scale lets lay participants test interventions in context and understand spatial relationships beyond static representations. This situated setting supports design judgment because interventions can be assessed embodied against the existing spatial configuration and perceived qualities of the place (SCHÖN 2017). Within this environment, first-person navigation is central, particularly because urban and landscape sites operate at relatively large scales. Moving through the scene across key areas helps participants develop spatial understanding of distance, adjacency, and scale, and supports checking of interventions in context (LYNCH 1960).

Building on this spatial grounding, hand tracking can further strengthen the simulator's suitability for participatory design by lowering the interaction barrier and increasing enjoyment. In this pilot, participants with no VR experience and no spatial design experience were still able to produce spatial interventions in Level 3 and described the process as enjoyable. Hand tracking aligns interaction with familiar embodied skills, shifting attention from memorizing mappings to expressing spatial intent. In this sense, "fun" supports sustained engagement and encourages non-experts to explore, intervene, and iterate within the virtual environment.

Earlier gesture-based systems in design contexts often treat gestures as a general-purpose alternative for selection and manipulation (MASUROVSKY et al. 2020). This work differs by operationalizing hand tracking as a comprehensive workflow for site-scale urban and landscape design, combining first-person navigation with hand-based interaction through a tailored design toolkit for contextual editing. The contribution lies in the interaction structure that links toolkit-based interventions and in-situ evaluation through repeated movement and iteration, rather than proposing gestures as a standalone input technique.

This formative pilot has clear limitations. First, hand tracking still showed occasional accuracy and sensitivity issues, indicating scope to improve gesture design and interaction stability. Second, the sample was small (N = 5). As an initial step, this pilot primarily assessed the simulator from a usability perspective. Future work should run larger, structured studies to confirm usability and to test whether the simulator effectively supports participatory design outcomes across users spanning different levels of VR and spatial design experience.

In summary, this paper introduced a VR simulator that operationalizes a hand-tracking interaction framework for urban and landscape design through toolkit-based interventions and contextual editing. The formative pilot suggests the system can support urban and landscape design activity in a first-person, human-scale environment, while revealing practical trade-offs between hand tracking and controllers that inform interaction design. Overall, the pilot points to hand tracking as a potentially accessible input modality for lay participants, supporting task completion and iterative exploration in participatory spatial design.

References

BUCKINGHAM, G. (2021), Hand Tracking for Immersive Virtual Reality: Opportunities and Challenges. Frontiers in Virtual Reality, 2, 728461. doi:10.3389/frvir.2021.728461.

CHOWDHURY, S. & SCHNABEL, M. A. (2019), Laypeople's Collaborative Immersive Virtual Reality Design Discourse in Neighborhood Design. Frontiers in Robotics and AI, 6, 97. doi:10.3389/frobt.2019.00097.

DOURISH, P. (2001), Where the action is: The foundations of embodied interaction. MIT Press.

LYNCH, K. (1960), The image of the city. MIT Press.

MALININ, L. H. (2016), Creative Practices Embodied, Embedded, and Enacted in Architectural Settings: Toward an Ecological Model of Creativity. Frontiers in Psychology, 6. doi:10.3389/fpsyg.2015.01978.

MASUROVSKY, A., CHOJECKI, P., RUNDE, D., LAFCI, M., PRZEWOZNY, D. & GAEBLER, M. (2020), Controller-Free Hand Tracking for Grab-and-Place Tasks in Immersive Virtual Reality: Design Elements and Their Empirical Study. Multimodal Technologies and Interaction, 4 (4), 91. doi:10.3390/mti4040091.

PORTMAN, M. E., NATAPOV, A. & FISHER-GEWIRTZMAN, D. (2015), To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. Computers, Environment and Urban Systems, 54, 376384. doi:10.1016/j.compenvurbsys.2015.05.001.

SANDERS, E. B.-N. (2002), From user-centered to participatory design approaches. In: J. Frascara (Ed.), Design and the Social Sciences (18-25). CRC Press/Taylor & Francis. doi:10.1201/9780203301302-ch8.

SCHÖN, D. A. (2017), The Reflective Practitioner. Routledge. doi:10.4324/9781315237473.

VAN LEEUWEN, J. P., HERMANS, K., QUANJER, A. J., JYLHÄ, A. & NIJMAN, H. (2018), Using Virtual Reality to Increase Civic Participation in Designing Public Spaces. In: BOUZAS-LORENZO, R. & CERNADAS RAMOS, A. (Eds.), Proceedings of the 18th European Conference on Digital Government (ECDG 2018) (pp. 230-239). Academic Conferences and Publishing International Limited.