Virtual Reality
https://doi.org/10.1007/s10055-022-00735-2

ORIGINAL ARTICLE

A hybrid 2D–3D tangible interface combining a smartphone and controller for virtual reality

Li Zhang[1] · Weiping He[1] · Huidong Bai[2] · Qianyuan Zou[2] · Shuxia Wang[1] · Mark Billinghurst[2]

Received: 21 June 2022 / Accepted: 2 December 2022

Abstract
Virtual reality (VR) controllers are widely used for 3D virtual object selection and manipulation in immersive virtual worlds, while touchscreen-based devices like smartphones or tablets provide precise 2D tangible input. However, VR controllers and touchscreens are used separately in most cases. This research physically integrates a VR controller and a smartphone to create a hybrid 2D–3D tangible interface for VR interactions, combining the strength of both devices. The hybrid interface inherits physical buttons, 3D tracking, and spatial input from the VR controller while having tangible feedback, 2D precise input, and content display from the smartphone's touchscreen. We review the capabilities of VR controllers and smartphones to summarize design principles and then present a design space with nine typical interaction paradigms for the hybrid interface. We developed an interactive prototype and three application modes to demonstrate the combination of individual interaction paradigms in various VR scenarios. We conducted a formal user study through a guided walkthrough to evaluate the usability of the hybrid interface. The results were positive, with participants reporting above-average usability and rating the system as excellent on four out of six user experience questionnaire scales. We also described two use cases to demonstrate the potential of the hybrid interface.

Keywords Hybrid interface · Virtual reality · Interaction paradigm · Tangible interface · Smartphone · Controller

1 Introduction

Interactions in the virtual worlds have always been an essential topic of virtual reality (VR), and rich interactions within VR have enhanced the user experience and benefited many applications. Handheld controllers are provided with commercial head-mounted displays (HMDs), becoming one of the most commonly used tools for interactions in VR. They have significant advantages in providing portable and efficient 3D input compared to other interface devices. On the other hand, smartphones with touchscreen input and multimedia output have become essential daily-used devices. Compared with handheld controllers, a smartphone with additional tracking accessories can be capable of similar VR input (Hartmann and Vogel 2018; Matulic et al. 2021) and provide familiar physical interaction patterns with tactile feedback (Bai et al. 2021).

Controller and smartphone interfaces are often used separately, but a hybrid interface could be created by combining them, providing more potential for novel and intuitive human-computer interactions. A hybrid controller/smartphone interface is a typical asymmetric bimanual system. Thanks to the flexibility of hands-on simultaneous asymmetric interactions (Guiard 1988), such interfaces could improve system performance and enhance user experience (Cho and Wartell 2015; Huang et al. 2021). For example, prior research has shown the benefits of hybrid interfaces for

✉ Li Zhang
zhl93@mail.nwpu.edu.cn

✉ Weiping He
weiping@nwpu.edu.cn

Huidong Bai
huidong.bai@auckland.ac.nz

Qianyuan Zou
qzou984@aucklanduni.ac.nz

Shuxia Wang
shuxiaw@nwpu.edu.cn

Mark Billinghurst
mark.billinghurst@auckland.ac.nz

1 Cyber-Physical Interaction Lab, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an 710072, China

2 Auckland Bioengineering Institute, The University of Auckland, 70 Symonds St., Auckland 1010, New Zealand

Published online: 18 December 2022

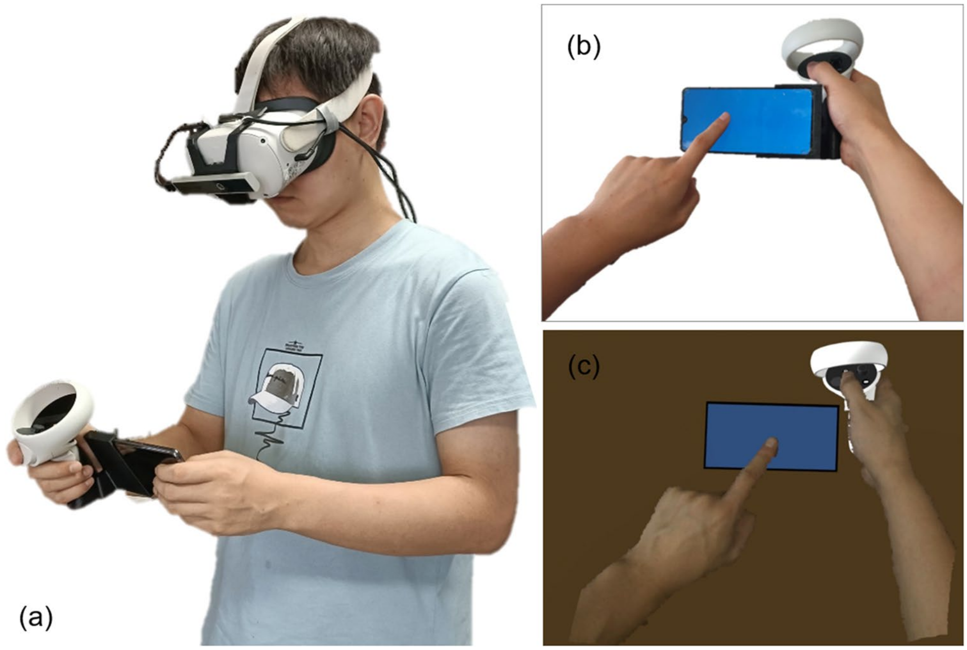


Fig. 1 Overall setup of the interface. **a** The overall VR system; **b** the interface in reality; **c** the interface in VR

supporting interactions in tabletop interfaces (Bragdon et al. 2011) and Augmented Reality (AR) (Kytö et al. 2018).

Similarly, earlier research has explored a range of hybrid interfaces by combining different interaction modalities in VR, such as hand gestures and a controller (Huang et al. 2021), a tablet and stylus (Billinghurst et al. 1997), and brain activity and eye gaze (Ma et al. 2018). However, there has yet to be a combination of a smartphone and handheld controller in VR. The lack of explicit design space and demonstrations of this combination limits smartphone-based touchscreen interactions in VR and potential advancements in many applications.

Inspired by the prior work on bringing a smartphone into VR (Bai et al. 2021; Zhang et al. 2020) and smartphone-based interactions in AR (Zhu and Grossman 2020), this research focuses on physically integrating a handheld controller and smartphone to create a hybrid VR interface that supports 2D–3D tangible interaction (see Fig. 1). Our prior idea focused on combining the devices based on the needs of customized functionalities (Zhang et al. 2021a, b, c), while the current interface design is driven by comparing the capabilities of VR controllers and smartphones toward VR interactions, respectively.

The hybrid interface inherits physical buttons, 3D tracking, and spatial input from the VR controller, and a touchscreen, tangible feedback, 2D precise input, and content display from the smartphone. The VR controller is dedicated to 3D positioning and trigger input, while the smartphone is used for 2D touchscreen input and multimedia output. Our novel solution provides an opportunity for rapid prototype design and demonstrations to explore an expanded design space for rich VR interactions.

We propose nine typical interaction paradigms (IPs) in VR with the hybrid interface by studying the design space and possible design principles. These IPs are composed to support more complex VR applications, including 3D model creation and manipulation, real walking with obstacle avoidance, and 3D text authoring. The implemented prototype showcases exemplary applications by demonstrating its unique advantages. A user study was also conducted to evaluate the interface usability and the user experience.

To our best knowledge, this is one of the first studies that integrate a handheld controller and smartphone to support hybrid 2D–3D tangible interactions in VR. Our research makes the following contributions:

1. We study design principles for hybrid 2D–3D tangible interfaces for VR that physically integrate a handheld controller and smartphone, grounded by literature and comparing their respective capabilities.
2. We discuss a design space and typical IPs based on the hybrid interface to provide 2D–3D tangible interactions in VR and implement a prototype system of the hybrid interface with three application modes by composing different IPs.
3. We present insights and findings of the hybrid interface from a user study regarding the integration effect of inherited interfaces and overall system usability.

2 Related works

Our work builds upon prior research in using smartphones in VR, VR tangible interactions, and VR hybrid interfaces. This section reports related works on these topics and highlights the differences from this work.

2.1 Smartphone-based interaction in VR

The research and applications on smartphone-based VR interactions have been growing recently. For example, smartphones have been used to provide touchscreen input without 3D tracking in VR (Liang et al. 2016; Mine et al. 2014; Gugenheimer et al. 2016). Users could rotate the phone with a gyroscope to support pan and zoom operations of virtual content or objects (Büschel et al. 2019; Vinayak et al. 2016). By combining these techniques, smartphones have also been used for 3D data manipulation (Issartel et al. 2017; Besançon et al. 2016).

A smartphone can be tracked in VR by attaching a tracking marker on the phone (Normand and McGuffin 2018; Biener et al. 2020; Matulic et al. 2021) and the HMD (Afonso et al. 2017), or connecting to a VR controller (Zhang et al. 2020), tracking sensors (Menzner et al. 2020; Zhu and Grossman 2020; Millette and McGuffin 2016), and hand tracking (Chang et al. 2018). The mobile SLAM technique can calculate 3D positions of the smartphone using an inside-out tracking approach (Babic et al. 2018; Ventura et al. 2014), but the tracking accuracy decreases with accumulated shift error. Some visual tracking techniques based on feature points are also widely applied in mobile interactions to provide stable tracking with a fiducial marker (e.g., Vuforia,[1] ARToolKit,[2] and Aruco[3]), but the cameras are occupied and the tracking area is limited.

Smartphone-based interactions are more commonly seen in AR because users can see their hands and the high-resolution phone screen wearing an AR HMD. In this way, a smartphone with tangible touchscreen input capability can allow rich interactions with the AR HMD (Dong et al. 2020; Lee and Chu 2018; Millette and McGuffin 2016; Zhu and Grossman 2020). However, it is still challenging to reproduce such smartphone-based interactions in VR without bringing the smartphone into VR.

2.2 Bringing a Smartphone into VR

An essential issue for bringing a smartphone into VR is that the VR HMD blocks the user from seeing the real phone and its screen content. However, the screen can be synchronously mirrored or captured with background service and shown in the virtual world, where the user wearing a VR HMD could operate the phone by interacting with the virtual screen (Takashina et al. 2018; Boustila et al. 2019; Kim and Kim 2016; Steed and Julier 2013). A second issue is that users cannot see their real hands in VR when they touch the screen, limiting their sense of presence (Grubert et al. 2018; Schwind et al. 2017). With an additional camera mounted on the front of the HMD, users could see a 2D hand image from a windowed region (Alaee et al. 2018) or a 3D point cloud (Desai et al. 2017). This setup enables users to access smartphone functionality within an immersive VR environment. However, the smartphone is not fully virtualized with a 3D model and cannot provide a high-resolution screen.

In our recent research (Bai et al. 2021; Zhang et al. 2020), a customized VR controller was attached to the phone to create a collocated virtual replica with real-time 3D tracking. A depth camera mounted on the front of the HMD provided a dense 3D hand point cloud with the help of a color-based filter. That work focused on using the smartphone in VR for the same activities as in the real world, such as sending text messages, and the attached controller was only used for 3D tracking. In contrast, in this paper, the controller not only provides 3D registration but also functions alongside the touchscreen to create a hybrid 2D–3D tangible interface in VR.

2.3 Tangible interfaces in VR

Tangible interfaces provide users with haptic feedback while enabling them to interact with digital information (Lindeman et al. 1999; Kim and Maher 2008). In a virtual environment, tangible interfaces can promote immersion and engagement (Wyeth 2008). When pointing, touching, or grasping virtual content, a collocated physical object can work as a proxy to provide coherent visual and tactile perception (Cheng et al. 2018; Hoppe et al. 2018).

The touchscreen is one of the most commonly used devices to provide physical force feedback with accurate pixel-level touchpoints. Touchscreens are widely used in VR scenarios that require precise input or selective input [e.g., sketching (Jetter et al. 2020; Drey et al. 2020) and text entry (Boustila et al. 2019; Kim and Kim 2016; Chen et al. 2019)]. The touchpoint on the 2D tablet screen can also be projected onto virtual objects to create a haptic illusion (Wang et al. 2019). However, the touchscreen is often provided by dedicated devices (electronic drawing board with a pen), which are hard to use in normal applications. Conversely, our hybrid interface provides an off-the-shelf solution for tangible spatial input, which is easy to use in many applications. The concomitant controller input further

1 https://developer.vuforia.com/.

2 http://www.hitl.washington.edu/artoolkit/.

3 http://www.uco.es/investiga/grupos/ava/node/26.

enlarges the design opportunities for smartphone-based tangible interaction.

2.4 Hybrid interfaces for VR

Hybrid interfaces are characterized by integrating familiar interaction paradigms with known design knowledge. Many studies have shown that hybrid interfaces can achieve better performance than a single device (de Haan et al. 2006). Regarding the interaction workflow, hybrid interfaces can be categorized as combinational, directional, and biased. The combinational hybrid interface integrates the capabilities from each side to create an entirely new device (Manuri and Piumatti 2015; Bornik et al. 2006; Sharmila 2021; Gao et al. 2020). The directional hybrid interface builds cross-device interactions when simultaneously operating two or more independent devices. For example, smartphones can provide spatial and touchscreen input when interacting with other devices (Millette and McGuffin 2016; Zhu and Grossman 2020; Schmidt et al. 2012). The biased hybrid interface tends to utilize other devices to enhance the capabilities and functions of the primary device. For example, smartphones are used to enhance controller-based VR interactions (Mohr et al. 2019; Hartmann and Vogel 2018). Our work explores the design space afforded by integrating a smartphone and a VR controller into a single device.

In terms of 2D touchscreen and 3D spatial input, they have been combined to benefit data visualization and understanding (Sommer et al. 2015; Drey et al. 2020; Büschel et al. 2019; Vinayak et al. 2016; Lee and Chu 2018). In contrast, this research focuses on synchronous operations with a combinational hybrid interface by integrating a VR controller and a smartphone. The interface is physically connected and self-contained, enabling 2D touchscreen and 3D spatial interactions with the virtual content. Although our prior work has demonstrated the combination (Zhang et al. 2021a) and potentials (Zhang et al. 2021b), in this paper, we further explore the design space, basic interaction paradigms, and application modes with the interface, as well as a user evaluation of the interface with a guided walkthrough.

3 Combining the controller and Smartphone

The hybrid interface aims to combine the specific properties of the VR controller and the smartphone. To do this, we first analyze and compare the capabilities of each interface and identify the design principles.

3.1 Capabilities of VR controllers

Typically, modern VR controllers are specifically designed to support interactions with a corresponding VR HMD. Since they can be vastly different from each other, we specify the VR controllers as the peripheral devices of mainstream VR HMDs, such as HTC Vive, Oculus Rift/Quest, and Microsoft Mixed Reality.

One benefit of VR controllers is that they can provide finger-based physical inputs with embedded triggers, buttons, touchpads, and joysticks. Another important characteristic of VR controllers is real-time spatial tracking at a high refresh rate. The physical inputs and tracking capability allow rich interactions with the virtual environment, including teleportation, selection, and object manipulation.

Limited by the purpose of VR controllers, the support for tangible interaction is not a primary design target. Although some VR controllers provide a touchpad for tangible input, the limited size of the panel cannot support the same large-scale touchscreen feedback as a smartphone. Another obvious limitation is that direct 3D operations in midair cannot achieve the same accurate and precise input obtained with 2D tangible devices (Wang et al. 2019).

3.2 Capabilities of smartphones

Touchscreen input on modern smartphones is familiar to many users. A huge advantage is that the high-definition screen and multi-touch recognition enable efficient and precise tangible 2D input with the touch panel while showing detailed images and texts. In addition, the tangibility of the smartphone is further applied in many scenarios. The familiar gestural input patterns are also used to provide interactions with the virtual content (Dong et al. 2020). Various embedded sensors provide ubiquitous connections with other devices and an exhaustive perception of the surrounding environment.

However, the lack of independent and robust tracking greatly limits the use of smartphones in both VR and AR. Since modern smartphones focus more on touchscreen-based gestural input, various physical buttons are no longer provided. Furthermore, people tend to hold the phone with one hand and use the other to operate the screen. Although one-handed operations are supported, the overall performance and usability are far worse than bimanual operations.

3.3 Design principles

VR controllers and smartphones have contrasting and complementary properties, as shown in Table 1. This motivates the combination of a VR controller and a smartphone to create a novel hybrid 2D–3D solution. Based on these investigations, we present design principles for the hybrid interface as discussed below.

(1) ***Two-hand operations***. When using a mobile phone for touchscreen interaction, it is common to hold the smartphone with one hand and use the other for touch input (Le

Table 1 Capabilities of VR controllers and smartphones (Yes: supported; Limited: partially supported; No: not supported)

	Items	VR Controller	Smartphone
Input	Spatial tracking	Yes	Limited
	Precise input	Limited	Yes
	Spatial input	Yes	No
	Virtual UI	No	Yes
Hardware	Touchscreen	Limited	Yes
	Physical Button	Yes	Limited
	Physical sensors	Limited	Yes
Output	Force feedback	Yes	No
	Content display	No	Limited
	Tangible feedback	Limited	Yes
Operation	Familiar interaction	No	Yes
	One-hand operation	Yes	Limited

et al. 2016). Nevertheless, with a hybrid interface, the hand holding the VR controller can also provide physical support for the smartphone.

(2) ***Phone for 2D; controller for 3D***. The VR controller features real-time spatial tracking with a high refresh rate and finger-based physical inputs. The smartphone provides virtual graphical user input and precise tangible 2D input. Thus, the smartphone should be used for precise and detailed 2D interactions, while the VR controller should be used for large-scale spatial input and coarse 3D interactions (Bai et al. 2017).

(3) ***Bidirectional remapping***. The smartphone and VR controller are both independent interfaces. The touch input on the smartphone and the spatial input with the VR controller can be directly remapped to each other, which means that operations with one device can be dynamically swapped and completed by the other or the hybrid interface. The phone screen content or the 3D virtual space should also allow a seamless transition to support bidirectional interactions (Zhu and Grossman 2020).

(4) ***Preserved interaction patterns***. Combining off-the-shelf devices with familiar IPs will shorten relearning and provide a more user-friendly interface. Instead of defining new IPs, one intuitive solution is to preserve the legacy integration patterns for both the smartphone and the VR controller.

4 A hybrid 2D–3D tangible interface

To systematically study the integration, we create a design space that integrates smartphone and controller design parameters. Different interaction modalities are provided by grouping various interaction metaphors. Based on the design space, we also present several typical IPs with the hybrid interface.

4.1 The design space

The design space includes three dimensions concerning the primary input device, controller function, and smartphone features. As shown in Table 2, we create a multi-dimensional morphological matrix with different values of the three dimensions extracted from the comparisons between smartphones and controllers. Each row with a combination of smartphone features represents an IP in the design space.

Primary input represents how the smartphone and the controller are used when interacting with the VR space. The primary input methods can be provided by the smartphone, controller, or both in a hybrid form. The handheld controller provides mandatory 3D tracking for the interface in all conditions. Its built-in triggers, buttons, and joysticks can also be used to provide various inputs. Moreover, the controller buttons can support supplementary interactions in a phone-centric design while users mainly focus on phone-based interactions. For smartphones, we provide five typical features to choose from (i.e., touchscreen swipe, touchscreen tap, content display, physical sensors, and familiar interaction patterns). These features are dependent in most cases so that several elements can be composed and used simultaneously for a specific IP.

4.2 Interaction paradigms

Some IPs with the hybrid interface can be retrieved from the morphological design space. As shown in Table 3, we present nine typical IPs by selecting different parameters, which are supposed to create unique user experiences that are different from current techniques. These IPs are demonstrated in Table 4. Although some IPs seem achievable with only the controller or the smartphone, the interaction can be realized differently, and the user experience can be significantly different.

IP1: ***Smartphone in VR***.

Introducing a full-featured smartphone to the virtual environment allows users to use the phone naturally without taking off the VR HMD (Bai et al. 2021; Zhang et al. 2020). The controller provides physical support for the smartphone and additional physical button input. In addition, an enlarged phone screen can be placed in the VR space, and users can interact with the content with the controller-based ray cast or hand gestures (Takashina et al. 2018).

IP2: ***Surface gestural input***.

Users are familiar with the interaction patterns of typical surface gestures with a smartphone from daily experience (e.g., dragging, tapping, swiping, and pinching). Some functions with the controller can be redirected to

Table 2 A design space of IPs with the hybrid 2D–3D tangible interface. Each row determines an IP with a primary input device, a VR controller function, and a combination of different smartphone features

Primary input	VR controller	Smartphone				
		Swipe	Tap	Content	Sensors	Familiarity
Smartphone	3D Tracking	□	□	□	□	□
	Buttons + tracking	□	□	□	□	□
Controller	3D Tracking	□	□	□	□	□
	Buttons + tracking	□	□	□	□	□
Both	Buttons + tracking	□	□	□	□	□

familiar surface gestures and benefit interaction modalities for VR (Menzner et al. 2020; Dong et al. 2020; Vanukuru et al. 2020; Matulic et al. 2021). In addition, handwriting input of characters and numbers is also helpful for text entry in VR.

IP3: ***Tangible virtual buttons***.

Given the limited built-in physical buttons on the controllers, the definition of smartphone-based virtual buttons could be a promising way. The physical buttons on the controller can also be remapped to the virtual buttons on the phone screen and vice versa. In addition to the familiar input patterns on the smartphone, tangible feedback can enhance the immersiveness and naturalness of VR interactions.

IP4: ***2D–3D content transfer***.

Content transfer between the smartphone and the VR space has become an interesting topic. The hybrid interface brings a unique opportunity for bidirectional content transfer between smartphones and the VR space (Biener et al. 2020; Zhu and Grossman 2020; Surale et al. 2019). The 2D content on the phone screen can be dragged into the spatial VR space to create a 3D presentation. Conversely, 3D objects in VR can also be transformed into 2D images on the phone screen when they are collided with the smartphone or selected by a ray emitted from the controller.

IP5: ***Onscreen tangible input***.

Onscreen tangible operation is often seen in video see-through mobile AR and VR applications. The phone screen is between the virtual space and the eyes so that users can simultaneously see and select user interface elements when operating on the panel (Surale et al. 2019). A similar process can also be introduced into VR with the hybrid interface. The phone screen can be virtually extended where the instructions and contextual information can be displayed beyond the screen region.

IP6: ***Augmented virtuality***.

Instead of seeing the physical world by mounting a camera on the VR HMD (Bergé et al. 2014), the smartphone camera can also capture the physical world from the controller position (Zhang et al. 2021c). With a hybrid interface, a user could customize the window and perspective to observe the physical surroundings in VR by mirroring the phone screen into the virtual environment. This Augmented Virtuality (AV) solution will help users confidently avoid physical obstacles and enable walking in the real world instead of teleportation or virtual walking (Bozgeyikli et al. 2016), which can enhance the user experience and decrease cybersickness.

IP7: ***Second virtual perspective***.

Given the hybrid interface, the phone screen can provide a second view of the virtual environment. Like a real camera, a second virtual camera view can allow users to

Table 3 Combinations of the primary input device, VR controller function, and smartphone features for each proposed IP with the hybrid interface. (☒: selected features)

No.	Interaction paradigms	Primary input	VR Controller	Smartphone				
				Swipe	Tap	Content	Sensors	Familiarity
IP 1	Smartphone in VR	Smartphone	3D Tracking	☒	☒	☒	☒	☒
IP 2	Surface gestural input	Smartphone	3D Tracking	☒				☒
IP 3	Tangible virtual buttons	Smartphone	3D Tracking	☒	☒	☒		☒
IP 4	2D–3D content transfer	Smartphone	Buttons + tracking	☒	☒	☒		☒
IP 5	Onscreen tangible input	Smartphone	Buttons + tracking	☒	☒			
IP 6	Augmented virtuality	Controller	3D Tracking			☒	☒	
IP 7	Second virtual perspective	Controller	Buttons + tracking			☒		☒
IP 8	Asymmetric operation	Both	Buttons + tracking	☒	☒	☒		☒
IP 9	Tangible spatial input	Both	Buttons + tracking	☒	☒			☒

naturally observe the virtual space from another perspective beyond the main viewpoint. This view can also be provided to observe the targets sideways in 3D VR space. Such a design corresponds to the normal habit of using a real smartphone, making users feel natural and familiar.

IP8: ***Asymmetric operation***.

Instead of two controllers, the asymmetric bimanual operation with a hybrid interface is enabled for synchronous or alternative 2D and 3D input (Bai et al. 2017). This provides an opportunity to enhance the operability and efficiency for virtual object manipulation (Kim and Park 2014; Wang et al. 2019). Furthermore, the relative spatial position of the smartphone and the controller can be swapped to provide different hand distributions according to the task. For example, the dominant hand can operate on the touchscreen for precise input, while the non-dominant hand can hold the controller for target selection.

IP9: ***Spatial tangible input***.

Regarding drawing and sketching in VR, this hybrid interface naturally meets all the requirements (Drey et al. 2020). The touchscreen offers pixel-level input and tangible feedback when drawing with fingers, while the controller stretches the 2D content on the phone screen into the 3D space. We believe the hybrid interface provides an instant-on solution for tangible spatial input with off-the-shelf devices, which helps users focus more on higher-level interaction design than the hardware.

5 Interactive prototype

Based on the design space and the IPs, we developed an interactive prototype of the hybrid interface and implemented three VR scenarios to demonstrate the application modes made by composing the IPs.

5.1 System implementation

5.1.1 Overall setup

As shown in Fig. 1, the prototype system contains a VR HMD with a controller, an RGB-Depth camera mounted on the front panel of the HMD, and a smartphone attached to the VR controller. We used an Oculus Quest 2[4] VR HMD since its resolution (1832×1920px per eye) supports a relatively clear perception of the phone screen content in VR. Compared with other mainstream VR HMDs, the built-in SLAM algorithm of the Quest 2 provides standalone room-scale tracking without external devices. The controller features rich physical buttons, proper size, and stable tracking, which is suitable for this research. Moreover, the alternative all-in-one design provides an easy deployment on the mobile platform. The Quest 2 HMD was connected to a PC with an Oculus Link cable and using Link Mode.

5.1.2 Hardware configuration and calibration

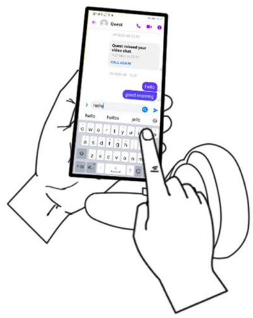
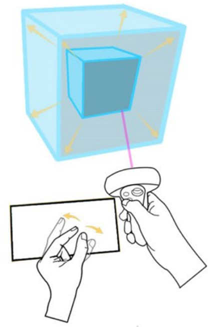
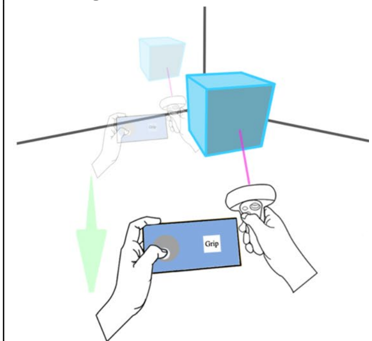
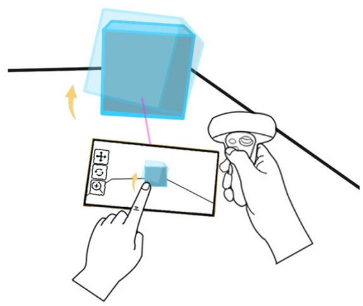
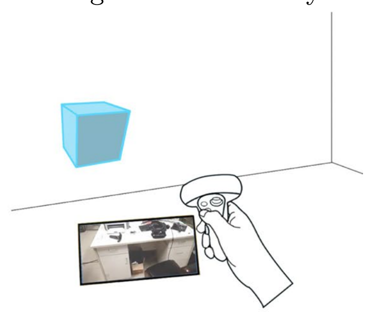
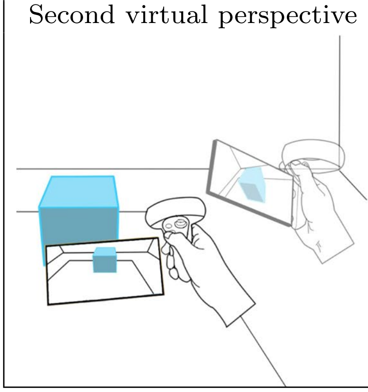
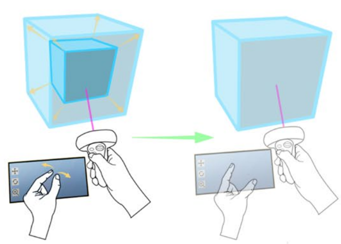
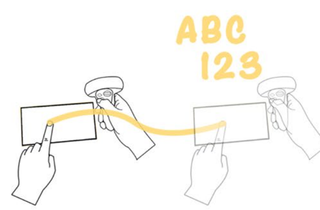
The phone was firmly connected to the VR controller by a 3D-printed structure. The controller handle was partly plugged into a hollow grip,[5] which does not affect the grasping of the controller. The smartphone was located and fixed in a case-like slot on the other side of the connector. The relative position between the controller and the smartphone was retrieved from the 3D connector model to render a collocated virtual smartphone in VR.

The RGB-Depth camera was horizontally installed in front of the HMD panel with a flexible thin-shelled 3D printing cover. Since the depth camera's field of view (FoV) is smaller than the HMD, the camera was vertically tilted down

[4] https://www.oculus.com/quest-2/.

[5] https://www.thingiverse.com/thing:4640517.

Table 4 Typical IPs proposed within the design space. Each cell represents an IP in a specific moment, although broader interaction concepts are included in the context

IP 1 Smartphone in VR 	IP 2 Surface gestural input 	IP 3 Tangible virtual buttons 
IP 4 2D-3D content transfer 	IP 5 Onscreen tangible input 	IP 6 Augmented virtuality 
IP 7 Second virtual perspective 	IP 8 Asymmetric operation 	IP 9 Tangible spatial input 

by 16 degrees to capture the hand input region. We also conducted a calibration procedure to align the camera and the hands with the corresponding virtual model and point cloud in VR (Bai et al. 2021).

5.1.3 Phone virtualization and hand segmentation

The VR controller provides six-DoF stable tracking for the phone through the connector, which only covered a small part of the edge and had no occlusion on the screen. We used the Scrcpy[6] Android mirroring software to wirelessly cast the phone screen to the window on the VR computer. The OBS Virtual camera software[7] was used to capture the window and stream it to the VR application as a live camera feed. The screen was finally rendered as a live texture on the 3D phone model surface.

[6] https://github.com/Genymobile/scrcpy/.

[7] https://obsproject.com/forum/resources/obs-virtualcam.949/.

An Intel SR305[8] camera was used to capture the dense point cloud of the near-field physical scene. We used an HSV-based filter to remove the points without the hand region. The filtering was defined in a shader and conducted in the final rendering procedure in the GPU so that the post-process had little effect on the frame rate. After the segmentation, only the skin-colored point cloud remained for visualization, and users could see their hands in the virtual environment in real-time.

5.1.4 Bidirectional communication

We built a TCP/IP communication channel as a background service to send commands and messages between the VR PC and the smartphone. This channel was independent of the screen-casting channel. A TCP/IP server was kept running on the PC and listening to new incoming clients. The smartphone and the VR system were connected to the server as independent clients. Commands and messages (e.g., touchpoint, selected object ID, and text) from a client were sent to the server and forwarded to the other after the connection was established. The TCP/IP communication mechanism supported synchronization between the smartphone and the controller in the virtual space.

5.1.5 Implementation and performance

The system was set up with a desktop PC (Intel Core i7-10700 CPU @ 2.9GHz, 16GB Memory, RTX 3070Ti GPU, Windows 10 Pro OS, 1GB Ethernet), and an Android smartphone (Huawei Mate 20, Android 10, 6G RAM, 6.53-inch planar HD screen with a resolution of 2244×1080@60Hz). They were connected to a local WiFi 6 network with a round trip time of about 6 ms. The phone screen was first mirrored to the VR PC with a resolution of 924×1920@60Hz, limited by the monitor resolution. The mirrored window was then cast to the VR system via OBS software as a live raw texture with a dropped frame rate of about 48 Hz. The screen mirroring occupied about 48 Mbps bandwidth with a latency of about 36 ms. The TCP/IP channel occupied less than 1 Mbps bandwidth with a latency of about 9 ms.

With the additional connector weight (36.6 g), the total weight of the hybrid interface was 374.7 g. The controller provided stable tracking for the interface at 60 Hz. The depth camera captured the physical scenario with a resolution of 640×480 @ 60Hz. The VR system was developed with Unity 3D Engine with a content update frequency of 98Hz and occupied 47% CPU and 72% GPU.

[8] https://www.intelrealsense.com/depth-camera-sr305/.

5.2 Application modes

With the hybrid interface, we further demonstrate three application modes composed of a combination of the IPs mentioned above. These modes cover most typical VR interactions, including smartphone-based interaction (mode 1), 3D object operation with asymmetric input (mode 2), and spatial 2D–3D input as an authoring tool (mode 3).

5.2.1 Mode 1: Smartphone-based interaction

A distinct advantage of the hybrid interface is using the smartphone in a virtual environment as naturally as in the real world. Users can answer phone calls, send messages, type texts, or browse multimedia content in VR without taking off the HMD (**IP1**). The phone provides full-featured multimedia, social communication, and interactive content for users with familiar interaction patterns. As shown in Fig. 2a, virtual content can be imported into the 3D VR space by dragging (**IP2**) the 2D picture of an object model into a specific region on the screen to instantiate it in the phone's current location (**IP4**). The model then becomes a three-dimensional object which can be interacted with in VR.

As shown in Fig. 2b, users can operate a virtual joystick on the screen to move around in 3D space and touch a virtual button to select the imported 3D model via a virtual laser beam (**IP3**). These buttons can be combined according to the situation and operation convenience. For example, the controller trigger is pressed to keep the virtual object selected, and the virtual joystick is used to navigate to a remote place virtually and vice versa (**IP8**).

Regarding body movement in VR, it might be more natural for users to walk instead of virtual teleportation. As shown in Fig. 2c, the phone camera can be opened to allow users to see the real world from the screen while walking around to avoid obstacles confidently (**IP6**).

5.2.2 Mode 2: 3D object operation with asymmetric input

After importing 3D models into the VR system, object manipulation is then required. As shown in Fig. 3, by pressing different triggers on the controller, participants could choose between controller-based operation and phone-based operation or both (**IP8**). Users can press the index trigger to enable raycasting-based object selection and controller-based collocated 3D operation. At the same time, the hand trigger can be pressed to allow additional touchscreen-based rotation, translation, or scaling based on swipe gesture recognition (**IP2**). Icons and virtual buttons are displayed on the smartphone screen for users to choose input and operation methods (**IP3**). The visualized virtual hands help users to interact with the virtual phone tactically and naturally.

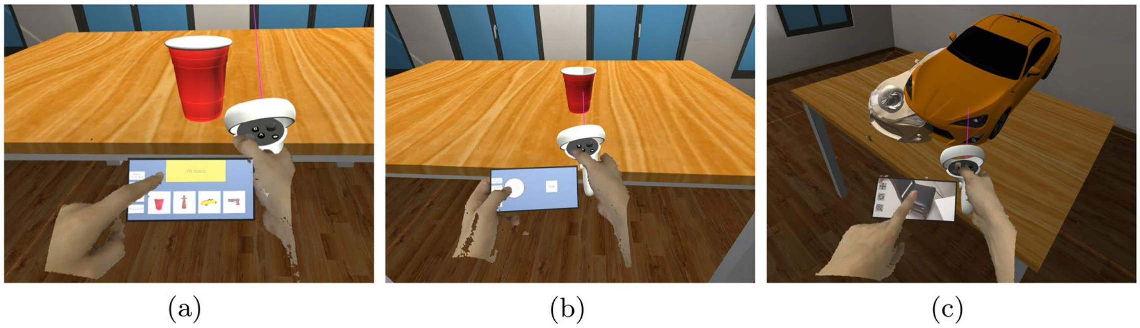


Fig. 2 **a** Putting the corresponding model into 3D space. **b** The virtual joystick is used for walking, and the controller trigger is pressed for object selection and manipulation. **c** Walking physically while observing the phone window

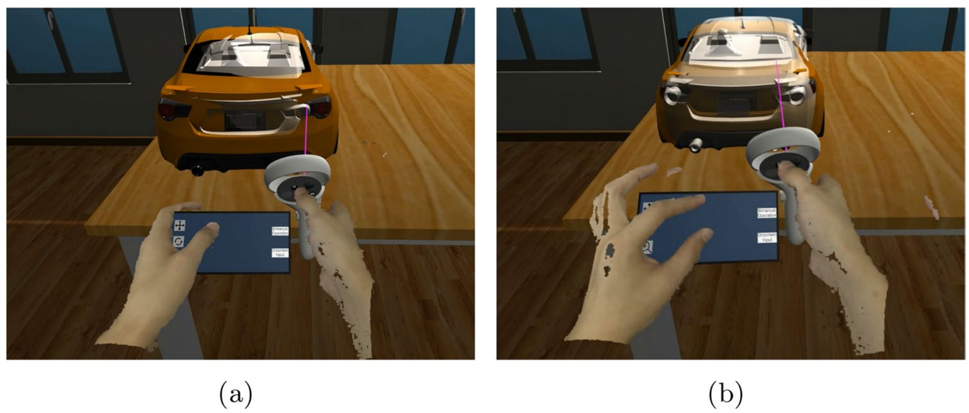


Fig. 3 Operations with the controller and touchscreen at the same time. The touchscreen provides gesture-based **a** translation, rotation, and **b** scaling input

Fig. 4 **a** Onscreen operations while seeing the virtual content on the smartphone display. **b** The second perspective provided by moving the phone screen

In VR, users can move the phone up between their eyes and virtual objects and activate a phone-bundled wide-view virtual camera to observe the VR scenario from the screen (Fig. 4a). The captured content is rendered on the screen from the phone's perspective (**IP5**), and the above input modalities are reserved. During the operation, the virtual camera lets the user perceive local details and view occluded content from the additional observation window (**IP7**). As shown in Fig. 4b, the virtual handheld camera can be flexible in the 3D space and serve as a reference window for a side view or a near-field window for local details so that users can avoid frequent head movement.

5.2.3 Mode 3: Spatial 2D–3D input as an authoring tool

The smartphone offers tangible feedback and functions as a touchpad to type words and sentences or draw lines and characters. Meanwhile, the controller provides live 3D position information to place the screen inputs into the virtual space. As shown in Fig. 5a, the smartphone keyboard provides familiar text entry support in VR by mirroring the

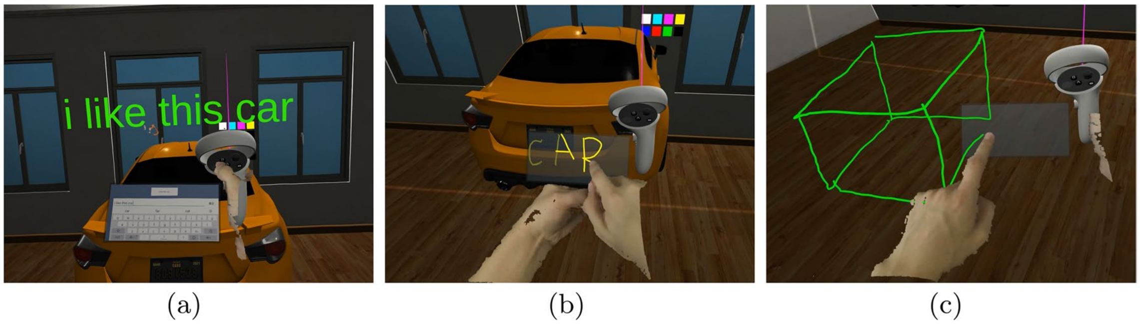


Fig. 5 **a** Converting text input on the smartphone to 3D virtual environment. **b** Writing comments on the 3D model with a different color. **c** Drawing a cube wireframe in VR

high-definition phone screen (**IP1**), and these words and sentences can be easily placed in the 3D environment (**IP4**). Thus, 3D text can be added by typing or drawing on the phone screen.

For the hybrid interface, 3D midair input can be decomposed into a 3D position and a temporary 2D surface (**IP9**). The virtual hands rendered in a point cloud format show the corresponding fingertips. Users can hold the controller and move their fingers on the touchscreen for short line segments and characters with real-time tangible feedback (see Fig. 5b). Thus, they could move the interface to different positions and repeat the above procedure to draw characters in 3D space.

Authoring tools are enabled for users to format the text in VR with the controller (**IP8**). For example, two buttons on the controller are used to adjust the font size, and a ray-cast-based color selection on a virtual palette is activated to change the text color. As shown in Fig. 5c, long line segments or large sketches are also supported. Users can touch the screen and move the controller simultaneously to stretch the touchpoints in VR.

6 User evaluation

We conducted a user study with our system to collect subjective feedback on the hybrid 2D–3D interface's usability and user experience. The study focused on showing these IPs with the interface rather than exhaustively evaluating the system's performance.

6.1 Participants and procedure

We recruited fifteen participants (six females, nine males) aged 20-30 years (Mean=23.7, SD=1.98) for the study. They had different VR experiences (one was well-experienced, nine were familiar, and five had no experience), but none had ever used a smartphone in VR. A short training session was conducted for those participants with little VR experience to help them get used to the VR system.

Before the user study, participants were introduced to the study design and the overall setup.

Participants were first shown the interface and different IPs. They were then given a guided walkthrough of the interactive prototype, allowing them to experience the individual paradigms individually. While experiencing each IP, they were asked to finish some tasks to learn about the essential features. The guided walkthrough lasted 30 minutes for each person.

Participants were then asked to complete a more complicated guided task with the prototype. Different interaction modalities were selected by touching virtual buttons on the phone screen. The task was divided into three stages corresponding to the three application modes, which include all individual IPs.

Stage 1. The first stage was observing the VR smartphone and dragging models into the virtual environment. After selecting a model with a virtual laser emitted from the controller, the participant needed to move it onto a virtual table about 3 meters away by virtually traveling with the virtual or physical joystick. Then, the phone camera was opened, and the participant must observe the screen and move another model to the table by physically walking and avoiding obstacles.

Stage 2. The second stage was to align two pairs of models with their copies. One model was aligned with synchronous or progressive touchscreen input and controller input, while the other was aligned by observing the virtual content from the touchscreen.

Stage 3. The final stage was text entry and spatial sketching. Characters were transferred from the phone keyboard into 3D space and formatted in different colors and sizes. The study ended with drawing a wireframe cube and adding handwriting labels.

After the user study, participants completed a subjective questionnaire and a short interview. We used the System Usability Scale (SUS) (Laugwitz et al. 2008) to measure the

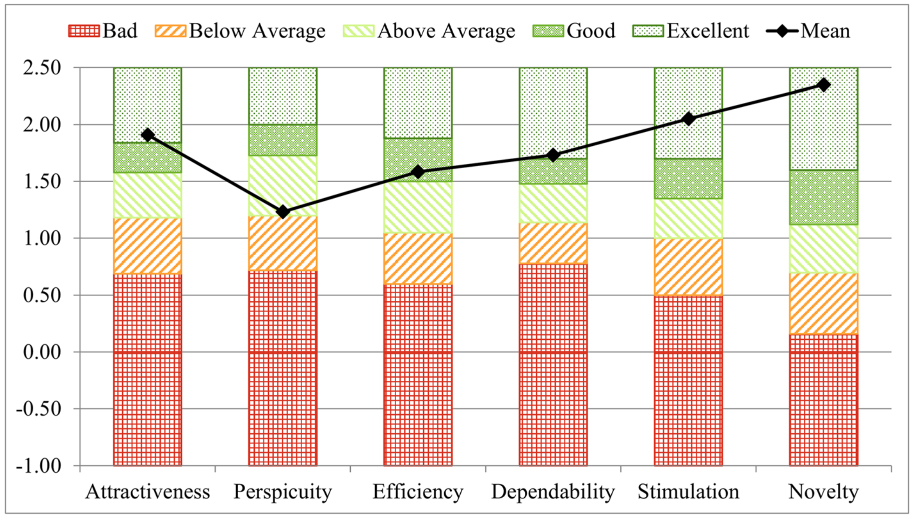


Fig. 6 Result of the UEQ survey (black square: mean; bars: benchmark) (Color figure online)

usability of the interface and the User Experience Questionnaire (UEQ) (Brooke 1996) to collect feedback on general user experience. Participants were also asked to rate each IP included in the system on a Likert scale from 1 (Poor) to 7 (Excellent). Detailed explanations of each IP were provided to help participants recall the corresponding experience. During the interview, participants were asked the following questions: (1) *What are your most and least favorite aspects of each interaction paradigm included in the system?* (2) *Do you have any comments on the integration of smartphones and controllers?* (3) *Under what conditions would you choose the smartphone and the controller, respectively?*

6.2 Result

6.2.1 System usability

The system usability was evaluated using the SUS questionnaire, consisting of 10 rating items with five response options (From Strongly Disagree to Strongly Agree). The overall SUS score ranges from 0 to 100, and a score of 68 or higher is regarded as above-average usability.

The SUS score of this interface was 70.8 (SD=16.8), which indicated that the usability of the interface was above average. Moreover, two participants rated the score very low (P4: 37.5, P9:27.5) because they could not see the content clearly when wearing glasses. After removing the two outliers, the SUS score increased to 76.7 (SD=7.9), which confirmed that users felt that the design and application mode of the hybrid interface was usable. Some participants reported that the system could be used in "*a lot of application scenarios*" (P4, P5). Although the interface was "*a bit heavy*" (P3, P4, P11) and "*required two hands*" (P3), it was "*attractive with good experience*" (P7, P8, P15). Besides, participants suggested to "*customize some settings as some interaction modalities were not the same as expected*" (P1, P5).

Table 5 The 5% confidence intervals for single items of the UEQ survey

Scale	Mean	Std. Dev.	Confidence	Confidence interval	
Attractiveness	1.911	0.617	0.312	1.599	2.223
Perspicuity	1.233	1.230	0.622	0.611	1.856
Efficiency	1.583	0.939	0.475	1.108	2.058
Dependability	1.733	0.821	0.415	1.318	2.149
Stimulation	2.050	0.751	0.380	1.670	2.430
Novelty	2.350	0.632	0.320	2.030	2.670

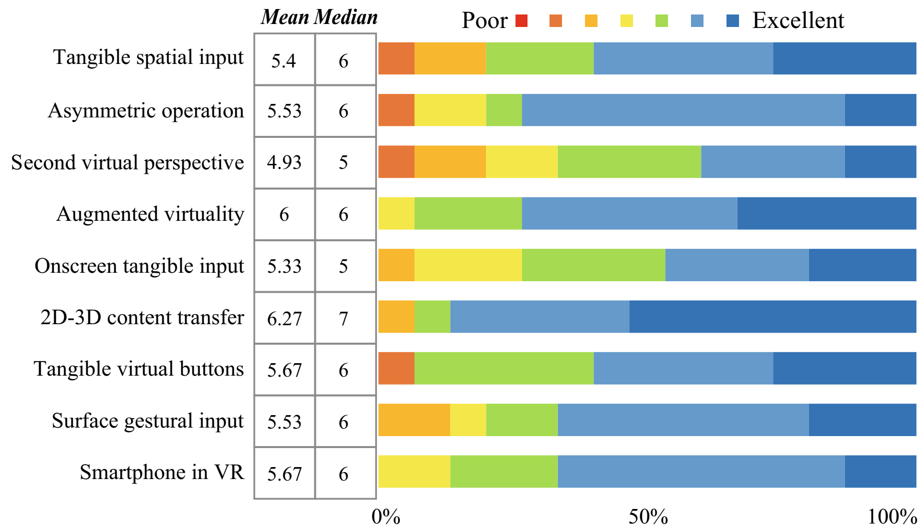
6.2.2 User experience

The UEQ has 26 contrasting attributes, each of which is evaluated on a 7-point Likert scale. With a UEQ data processing tool,[9] the items were rearranged to six scales and compared with the provided benchmark, as shown in Fig. 6 and Table 5. The result revealed that the hybrid interface was in the range of the 10% best results for the attractiveness, dependability, stimulation, and novelty scales. On the other hand, the hybrid interface ranged from top 25% to 50% on perspicuity and top 10% to 25% for efficiency.

Such a result was not unexpected because participants had never seen such a hybrid interface before. For example, some users said drawing 3D content on the phone screen "*was awesome*" (P2, P4, P5, P11) and a "*new interaction modality*" (P4, P12, P15). Seeing the real world through the phone screen "*helped me walk in reality with confidence*" (P1, P2, P6, P8, P12) and "*avoid obstacles and remove the chairs away*" (P2, P14). The smartphone and the controller were integrated on the physical structure and functions, making the hybrid interface reliable and consistent. The touchscreen

[9] https://www.ueq-online.org/.

Fig. 7 Likert rating percentage on individual IPs



2D input "*avoided shakes with the controller*" (P1, P6, P7, P10), "*provided additional tangible feedback*" (P3, P5, P15) with a "*familiar touchscreen input*" (P4, P6, P9, P14). The virtual hands were "*sometimes lost*" (P12), but "*helped me see my fingertip*" (P10).

However, participants were not very skilled with the hybrid interface and these IPs in the VR environment, even though they had an initial training session. For example, the synchronous operation was "*faster*" (P5, P8), but participants "*will only choose this style when experienced*" (P10). Instead, progressive operation helped participants "*focused on single operation*" (P2, P4, P9, P10). We provided a gallery of all IPs for each participant at once, which might have negatively affected perspicuity. Most participants did not get used to VR interactions and hybrid 2D–3D inputs, so the hybrid interface had limited benefits for virtual object manipulation.

6.2.3 Likert scale rating

Participants rated each IP very positively after the guided walkthrough of the interface and the application modes. As shown in Fig. 7, the means of the rating items were larger than 5, except for **IP7** (Second virtual perspective, *Mean* = 4.93). In addition, the medians of the rating items were 6 or 7, except for **IP7** (*Median* = 5) and **IP5** (Onscreen tangible input, *Median* = 5). This indicated that most of the proposed IPs were well-designed. The hybrid interface was affordable for a larger variety of interactions, which justified the integration of the hybrid interface and the design space.

The subjective interview also proved the result. Although the second perspective "*provided an additional view*" (P1, P10) to "*see subtle details*" (P3), it was "*not so useful*" (P2, P13, P14), and the phone screen was "*small*" (P6). The onscreen operation made participants "*feel like playing smartphone games*" (P1, P3, P9, P15), but they could already "*see the whole VR scenario*" (P4, P10, P12, P13), and the content on the phone screen was "*not so clear*" (P6).

The most popular IP we found from the result was a 2D–3D content transfer (**IP4**), which allowed participants to drag the 2D image of a model into the 3D space. Participants reported that "*such transfer was cool and interesting*" (P3, P4, P8, P10, P12). It was "*convenient*" (P2) to "*see and move the model in VR*" (P10, P11, P14) instead of on the phone screen.

6.2.4 Observations on user behaviors

There were some interesting findings when observing user behaviors. When navigating the space, the joystick was first provided for participants to walk virtually to another place. They were conservative and careful to hold out a hand and moved closer to the destination. However, participants became more aggressive in body movements when the smartphone's front camera was opened. They moved the phone to look around the real world and stepped forward confidently. When a virtual table appeared in front, they observed the phone screen to see the real world and check if they could physically step forward instead of touching the virtual table recklessly. When they found a real chair in the way from the camera view, they bypassed it or moved it aside with the other hand.

The controller and touchscreen provided asymmetric input support for aligning a model with its target. The controller was used to quickly move the model to its target location from the initial position. However, some participants occasionally used the phone screen to make additional pose adjustments in this procedure. When the model was roughly aligned, the model size was adjusted preferentially over translation and rotation. The controller and the smartphone were often used synchronously to refine the model posture. Although translation and rotation were both supported with each device, the controller was focused on position adjustment, and the touchscreen was mainly used for pose refinement. This indicated the function allocation of the hybrid interface for precise virtual object manipulation.

Compared with the midair floating operation, the phone screen provided "*tangible feedback*" (P2, P3, P7, P9) with fingertip input and "*efficient rotation and scaling input*" (P3, P6, P11, P12, P14). The controller was mainly used in "*selection*" (P3), "*near-filed object manipulation*" (P9, P11, P13), "*large-scale movement*" (P4, P10, P14), "*fast but not precise movement*" (P2, P5), with "*physical feedback*" (P1, P12). In contrast, the smartphone was mainly used in content display, "*precise operation and adjustment*" (P4, P5), and "*far object manipulation*" (P9). Fortunately, "*they were well integrated to swap between each other*" (P12, P14) with "*fluent operations without obvious delay*" (P7, P13).

7 Discussion

This paper illustrates a smartphone and VR controller combination in a hybrid 2D–3D tangible interface for VR. The work demonstrates several unique interaction paradigms with the interface and showcases integrated application modes. This section discusses the design and implementation of the interface, interaction paradigms, and findings from the user study.

The smartphone and VR controller are capable of rich interactions with the VR environment, which provides the foundation for their integration on both the structure and function. The design principles and the design space were grounded by comparing smartphones and controllers regarding their capabilities. However, the comparison was general, and the result may not apply to all types of smartphones and VR controllers. Each dimension of the design space only chose the most typical features of smartphones and controllers. When integrating a specific smartphone and a controller, the design space can be slightly adjusted to add unique features (e.g., foldable screen, voice input, and fingerprint identification).

Based on the design space, we implemented nine typical IPs, which could be assembled to form a more complicated scenario. These new paradigms demonstrated the potential of the hybrid interface. However, some IPs (**IP5**, **IP7**) were not as efficient as expected. This means that the interactions with the combined hybrid interface toward VR are possibly totally different from existing interfaces.

Although some IPs can be realized with only a controller or a smartphone, the user experience is different. For example, geometry drawing (**IP9**) and text input (**IP1**) can also be achieved solely with a controller. However, the spatial sketching provided by a controller must be performed in 3D midair without tangible feedback on the fingers, leading to a limited perception of the depth. The 3D drawing is composed of touchscreen gestures and controller movement with this interface. The tangible perception of the fingers could enhance stability and the sense of control (Wang et al. 2021). Similarly, controller-based character selection on a virtual keyboard typically provides text entry in VR. Selecting characters in the air is not as natural and realistic as typing on a smartphone. These alternative interaction modalities allow developers to enhance the system experience and improve usability in different applications.

We provide an off-the-shelf solution to combine a smartphone and a controller in software and hardware, which can be easily expanded to the mobile VR platform. The interface features asymmetric bimanual operations for both 2D and 3D input devices. With virtual hand rendering and realistic smartphones, the prototype allows a variety of IPs and applications. The user study results also showed above-average system usability and positive user experiences. However, we admit that the setup is a bit clunky because of the weight and the additional depth camera, which could be removed when the virtual hands are unnecessary (e.g., **IP6** and **IP7**).

It seems that users could interact with a virtual smartphone in VR and only a rigid object in the real world. They could still get the same visual and tactile feedback, and the touch input could be obtained from finger detection in VR. However, the capabilities of touchscreens (e.g., high-precision input, high refresh rate, and accurate gesture detection) determine that the interactions provided by a touchscreen are challenging to be easily simulated with a quick and straightforward setup. The accuracy of current hand detection technology is far from meeting the needs of touchscreen interaction, which requires contact detection, submillimeter-level fingertip detection, multi-fingertip tracking and distinction, etc. The simulation-based method cannot achieve a similar user experience. In contrast, using the touchscreen is an easy approach for natural and efficient interactions with off-the-shelf devices and a quick setup.

Although the user study result is exciting and positive, the research has some limitations. The limited focus view of the depth camera constrained the region for capturing the hands. The lighting condition in the real world could also significantly affect the hand segmentation result. The prototype required the user to operate the controller with the right hand and the phone screen with the left hand. However, some users are left-hand dominant, so the right hand might be more efficient in operating the smartphone. More in-depth research on hand asymmetry can be conducted in the future. The user study might also be limited in the range of tasks studied, the types of data collected, and the time spent in the VR system.

8 Use cases with the Interface

The interactive prototype was evaluated with a guided walkthrough of some tasks concerning the research target. The experimental tasks cascaded these IPs to form a series of tasks from smartphone-based 3D model creation and manipulation to asymmetric bimanual operation and ended with authoring 3D comments. This section presents two use cases where the hybrid 2D–3D tangible interface can enhance the system performance and improve the user experience.

8.1 SWAI: Simultaneous walking and asymmetric input

Current VR navigation techniques, like teleportation and smooth viewpoint movement, allow users to walk in VR without moving their legs. However, the inconsistent sensation of acceleration, orientation, and proprioception could break the feeling of presence (Usoh et al. 1999) and lead to cybersickness and fatigue (LaViola 2000). This can be addressed by opening a window to the real world with an additional camera. Nevertheless, the camera is often fixed to HMD, and a large window of the real environment will break the immersive VR experience. A handheld camera with a limited viewport size will be a flexible solution to resolve this problem, especially when viewing local details of physical objects.

This hybrid interface lets users see the real world from the smartphone screen by opening the back camera. They can reach a destination by physically walking instead of teleportation and simultaneously interact with virtual content with both controller-based 3D and touchscreen-based 2D input. Thus, users can select a virtual object in an initial position by pressing a controller trigger and physically walking to another place to put down the object. Meanwhile, they can change the object pose or provide additional relative movement to the controller with touchscreen gestures. The screen size is not too large to break the immersive experience, and the camera's focal distance can be adjusted to obtain a wider view or see further objects.

Our prior research has demonstrated the scenario and collected user feedback (Zhang et al. 2021b). When participants walked with the interface, they tended to bypass obstacles or directly move them aside, becoming more confident about moving in the VR environment. They reported little influence of the window presenting the physical scene on the VR experience. They also appreciated the integration of the controller and smartphone. The controller was mainly used for large-scale manipulation to ensure efficiency, while the touchscreen input was used for additional zooming and rotation and precise adjustment in the final refinement stage.

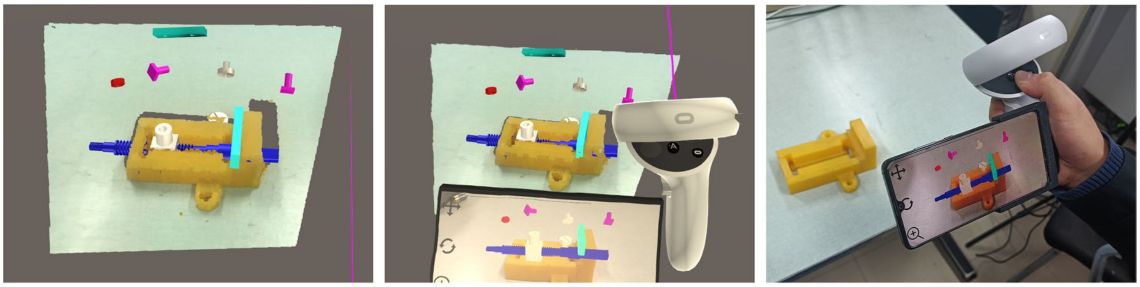
8.2 ARinVR: Bringing mobile AR into VR

Mobile AR allows users to observe the real world overlapped with 3D virtual content on smartphones or tablets. However, the limited screen size and lack of rich input modalities prevent users from efficiently interacting with the virtual content as wearing VR HMDs, especially in scenarios that require frequent interactions with models and menus (e.g., product assembly training and in-suit 3D sketching). A fiducial marker or a tracking sensor must be used for stable 3D positioning, and a hand must be occupied when holding the display. On the other hand, VR creates a purely virtual environment that allows users to concentrate on the task without being distracted by the surroundings. More authoring tools can be used for 3D interactions with bimanual input using 6-DOF controllers or gestures.

Therefore, a better solution for mobile AR interaction might be the combination of AR presentation and VR interaction with this interface (Zhang et al. 2021c). The physical scenario can be reconstructed with a 3D sparse point cloud and presented in the VR space as a reference. So users can wear a wide field-of-view VR HMD and leverage the VR-specific interaction tools and touchscreen to interact with the virtual content. After the task is finished in VR, users can directly observe the overall AR effect on the smartphone without taking off the VR HMD. We provide an example of AR assembly with this interface to illustrate more details.

As shown in Fig. 8a, the task is to assemble virtual models on an existing physical base. The physical base is scanned by the HMD-mounted depth camera and presented in VR in a colored point cloud format. Thus, users can manipulate virtual models with the controller for large-scale and coarse 6-DOF movement and adjust the target pose with precise touchscreen refinement (see Fig. 8b). This procedure is much more efficient and immersive than manipulating virtual objects with smartphone movement or hand gestures in AR.

Moreover, the point cloud presentation of the physical base allows users to perceive the AR effect in the VR environment. The AR rendering effect can also benefit from the occlusion between the point cloud base and virtual objects. As shown in Fig. 8c, the interface provides real-time 3D tracking for the smartphone and synchronizes the VR content to the mobile AR application. In addition to touchscreen input, the smartphone enables an AR presentation of the ongoing task by mirroring the screen to VR. Therefore, users can quickly switch between the mobile AR presentation space and the VR task space, enhancing the performance and interaction efficiency in physical AR tasks with a more enjoyable and immersive user experience.



(a) (b) (c)

Fig. 8 **a** The physical base is scanned by the HMD-mounted depth camera and presented in VR with a point cloud format. **b** The VR controller and the touchscreen are used to assemble virtual parts to the base. **c** The mobile AR screen is activated and mirrored to VR for the user to observe the AR effect in VR without taking off the HMD (Color figure online)

9 Conclusion and future work

This work presented a hybrid 2D–3D tangible interface created by combining a smartphone and a VR handheld controller in a VR environment, enabling simultaneous 2D and 3D interactions. We summarized the design principles and described a design space for creating interaction paradigms with the hybrid interface. An interactive prototype demonstrated nine typical IPs and three application modes. The user study confirmed the integration and the usability of the hybrid interface in a guided walkthrough. The interface can be reproduced quickly with off-the-shelf hardware and has the potential to be widely used in a variety of VR and AR applications. There is more research to do, but these preliminary results provide a promising start to using more hybrid smartphone/controller interfaces in VR experiences.

In the future, we would like to package the project as an open SDK for researchers to develop modular applications by composing different IPs. Furthermore, the physical settings of the system could be customized to lower the weight and enhance the performance, according to the design target. For instance, the depth camera can be removed when providing asymmetric virtual object manipulation, as users are mainly focused on the task rather than the interface.

We would also like to explore applications of the hybrid 2D–3D tangible interface in AR. The latest Passthrough support[10] of the HMD used in this work makes it easier to use this interface in the video see-through AR environment. For instance, the hybrid interface could provide precise target selection in a pinpointing application such as a CAD system.

Furthermore, we will create more unique interaction modalities based on current work. More application scenarios will be developed in our future work, including, (1) merging controller-based modalities into a smartphone, (2) controller-based interaction with a mirrored large phone screen in a VR collaboration task, and (3) precise virtual object manipulation with controller-based spatial input and touchscreen-based 2D input.

Author contributions LZ was involved in the conceptualization, methodology, validation, software, formal analysis, data Curation, writing—original draft and visualization. WH contributed to the conceptualization, resources, writing—review and editing, supervision and funding acquisition. HB assisted in the conceptualization, methodology, validation, writing—review and editing. QZ helped in writing—original draft and visualization. SW contributed to the resources, writing—review and editing and funding acquisition. MB contributed to the conceptualization, writing—review and editing and supervision.

Funding This work was partially supported by the National Key R &D Program of China (Grant Nos. 2019YFB1703800, 2021YFB1714900, 2021YFB1716200, 2020YFB1712503), the Programme of Introducing Talents of Discipline to Universities (111 Project), China (Grant No. B13044), the Fundamental Research Funds for the Central Universities, NPU (Grant No. 3102020gxb003).

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval Approval of all ethical and experimental procedures and protocols was granted by the Medical and Animal Care Ethics Committee, Northwestern Polytechnical University (No. 202202048).

References

Afonso L, Dias P, Ferreira C, Santos BS (2017) Effect of hand-avatar in a selection task using a tablet as input device in an immersive virtual environment. In: 2017 IEEE symposium on 3D user interfaces (3DUI). IEEE, pp 247–248

Alaee G, Deasi AP, Pena-Castillo L, Brown E, Meruvia-Pastor O (2018) A user study on augmented virtuality using depth sensing cameras for near-range awareness in immersive VR. In: IEEE VR's 4th workshop on everyday virtual reality (WEVR 2018), vol 10

[10] https://developer.oculus.com/experimental/passthrough-api/.

Babic T, Reiterer H, Haller M (2018) Pocket6: A 6dof controller based on a simple smartphone application. In: Proceedings of the symposium on spatial user interaction, pp 2–10

Bai H, Nassani A, Ens B, Billinghurst M (2017) Asymmetric bimanual interaction for mobile virtual reality. In: ICAT-EGVE, pp 83–86

Bai H, Zhang L, Yang J, Billinghurst M (2021) Bringing full-featured mobile phone interaction into virtual reality. Comput Graph 97:42–53

Bergé L-P, Serrano M, Perelman G, Dubois E (2014) Exploring smartphone-based interaction with overview+detail interfaces on 3d public displays. In: Proceedings of the 16th international conference on human-computer interaction with mobile devices & services. MobileHCI '14. Association for Computing Machinery, New York, pp 125–134. https://doi.org/10.1145/2628363.2628374

Besançon L, Issartel P, Ammi M, Isenberg T (2016) Hybrid tactile/tangible interaction for 3d data exploration. IEEE Trans Visual Comput Graph 23(1):881–890

Biener V, Schneider D, Gesslein T, Otte A, Kuth B, Kristensson PO, Ofek E, Pahud M, Grubert J (2020) Breaking the screen: Interaction across touchscreen boundaries in virtual reality for mobile knowledge workers. IEEE Trans Visual Comput Graph 26(12):3490–3502. https://doi.org/10.1109/TVCG.2020.3023567

Billinghurst M, Baldis S, Matheson L, Philips M (1997) 3d palette: A virtual reality content creation tool. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology. VRST '97, pp. 155–156. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/261135.261163

Bornik A, Beichel R, Kruijff E, Reitinger B, Schmalstieg D (2006) A hybrid user interface for manipulation of volumetric medical data. In: 3D user interfaces (3DUI'06). IEEE, pp 29–36

Boustila S, Guégan T, Takashima K, Kitamura Y (2019) Text typing in VR using smartphones touchscreen and HMD. In: 2019 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, pp 860–861

Bozgeyikli E, Raij A, Katkoori S, Dubey R (2016) Point & teleport locomotion technique for virtual reality. In: Proceedings of the 2016 annual symposium on computer-human interaction in play. CHI PLAY '16. Association for Computing Machinery, New York, pp 205–216. https://doi.org/10.1145/2967934.2968105

Bragdon A, DeLine R, Hinckley K, Morris MR (2011) Code space: Touch + air gesture hybrid interactions for supporting developer meetings. In: Proceedings of the ACM international conference on interactive tabletops and surfaces. ITS '11. Association for Computing Machinery, New York, pp 212–221. https://doi.org/10.1145/2076354.2076393

Brooke J et al (1996) Sus-a quick and dirty usability scale. Usability evaluation in industry 189(194):4–7

Büschel W, Mitschick A, Meyer T, Dachselt R (2019) Investigating smartphone-based pan and zoom in 3d data spaces in augmented reality. In: Proceedings of the 21st International conference on human-computer interaction with mobile devices and services. MobileHCI '19. Association for Computing Machinery, New York. https://doi.org/10.1145/3338286.3340113

Chang Y-K, Huang J-W, Chen C-H, Chen C-W, Peng J-W, Hu M-C, Yao C-Y, Chu H-K (2018) A lightweight and efficient system for tracking handheld objects in virtual reality. In: Proceedings of the 24th ACM symposium on virtual reality software and technology, pp 1–2

Cheng L-P, Chang L, Marwecki S, Baudisch P (2018) Iturk: Turning passive haptics into active haptics by making users reconfigure props in virtual reality. In: Extended abstracts of the 2018 CHI conference on human factors in computing systems. CHI EA '18, p. 1. Association for Computing Machinery, New York. https://doi.org/10.1145/3170427.3186482

Chen S, Wang J, Guerra S, Mittal N, Prakkamakul S (2019) Exploring word-gesture text entry techniques in virtual reality. In: Extended abstracts of the 2019 CHI conference on human factors in computing systems. CHI EA '19. Association for Computing Machinery, New York, pp 1–6. https://doi.org/10.1145/3290607.3312762

Cho I, Wartell Z (2015) Evaluation of a bimanual simultaneous 7D of interaction technique in virtual environments. In: 2015 IEEE Symposium on 3D User Interfaces (3DUI). IEEE, pp 133–136

de Haan G, Griffith EJ, Koutek M, Post FH (2006) Hybrid interfaces in ves: Intent and interaction. In: EGVE, pp 109–118. Citeseer

Desai AP, Pena-Castillo L, Meruvia-Pastor O (2017) A window to your smartphone: exploring interaction and communication in immersive VR with augmented virtuality. In: 2017 14th conference on computer and robot vision (CRV), pp 217–224. IEEE

Dong Z, Piumsomboon T, Zhang J, Clark A, Bai H, Lindeman R (2020) A comparison of surface and motion user-defined gestures for mobile augmented reality. In: Extended abstracts of the 2020 CHI conference on human factors in computing systems, pp 1–8

Drey T, Gugenheimer J, Karlbauer J, Milo M, Rukzio E (2020) VRSketchIn: exploring the design space of pen and tablet interaction for 3d sketching in virtual reality. Association for Computing Machinery, New York, pp 1–14. https://doi.org/10.1145/3313831.3376628

Gao L, Bai H, Billinghurst M, Lindeman RW (2020) User behaviour analysis of mixed reality remote collaboration with a hybrid view interface. In: 32nd Australian conference on human-computer interaction. OzCHI '20. Association for Computing Machinery, New York, pp 629–638. https://doi.org/10.1145/3441000.3441038

Grubert J, Witzani L, Ofek E, Pahud M, Kranz M, Kristensson PO (2018) Effects of hand representations for typing in virtual reality. In: 2018 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, pp 151–158

Gugenheimer J, Dobbelstein D, Winkler C, Haas G, Rukzio E (2016) Facetouch: Enabling touch interaction in display fixed UIS for mobile virtual reality. In: Proceedings of the 29th annual symposium on user interface software and technology. UIST '16. Association for Computing Machinery, New York, pp 49–60. https://doi.org/10.1145/2984511.2984576

Guiard Y (1988) The kinematic chain as a model for human asymmetrical bimanual cooperation. Adv Psychol 55:205–228

Hartmann J, Vogel D (2018) An evaluation of mobile phone pointing in spatial augmented reality. Extended Abstracts of the 2018 CHI conference on human factors in computing systems. CHI EA '18. Association for Computing Machinery, New York, NY, USA, pp 1–6

Hoppe M, Knierim P, Kosch T, Funk M, Futami L, Schneegass S, Henze N, Schmidt A, Machulla T (2018) Vrhapticdrones: Providing haptics in virtual reality through quadcopters. In: Proceedings of the 17th international conference on mobile and ubiquitous multimedia. MUM 2018, pp 7–18. Association for Computing Machinery, New York. https://doi.org/10.1145/3282894.3282898

Huang Y-J, Liu K-Y, Lee S-S, Yeh I-C (2021) Evaluation of a hybrid of hand gesture and controller inputs in virtual reality. Int J Human-Comput Interact 37(2):169–180

Issartel P, Guéniat F, Isenberg T, Ammi M (2017) Analysis of locally coupled 3d manipulation mappings based on mobile device motion. Presence Teleop Virt 26(1):66–95

Jetter H-C, Rädle R, Feuchtner T, Anthes C, Friedl J, Klokmose CN (2020) "In VR, Everything is Possible!": Sketching and simulating spatially-aware interactive spaces in virtual reality, pp 1–16. Association for Computing Machinery, New York. https://doi.org/10.1145/3313831.3376652

Kim YR, Kim GJ (2016) Hovr-type: Smartphone as a typing interface in VR using hovering. In: Proceedings of the 22nd ACM conference on virtual reality software and technology. VRST '16, pp 333–334. Association for Computing Machinery, New York. https://doi.org/10.1145/2993369.2996330

Kim MJ, Maher ML (2008) The impact of tangible user interfaces on designers' spatial cognition. Human-Comput Interact 23(2):101–137

Kim H, Park J-H (2014) Duplicatespace: Enhancing operability of virtual 3d objects by asymmetric bimanual interaction. In: 2014 11th international conference on ubiquitous robots and ambient intelligence (URAI). IEEE, pp 324–327

Kytö M, Ens B, Piumsomboon T, Lee GA, Billinghurst M (2018) Pinpointing: precise head- and eye-based target selection for augmented reality. Association for Computing Machinery, New York, pp 1–14. https://doi.org/10.1145/3173574.3173655

Laugwitz B, Held T, Schrepp M (2008) Construction and evaluation of a user experience questionnaire. In: Symposium of the Austrian HCI and usability engineering group. Springer, pp 63–76

LaViola JJ Jr (2000) A discussion of cybersickness in virtual environments. ACM Sigchi Bull 32(1):47–56

Le HV, Mayer S, Wolf K, Henze N (2016) Finger placement and hand grasp during smartphone interaction. In: Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems. CHI EA '16. Association for Computing Machinery, New York, pp 2576–2584. https://doi.org/10.1145/2851581.2892462

Lee C-J, Chu H-K (2018) Dual-MR: Interaction with mixed reality using smartphones. In: Proceedings of the 24th ACM symposium on virtual reality software and technology. VRST '18. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3281505.3281618

Liang H-N, Shi Y, Lu F, Yang J, Papangelis K (2016) Vrmcontroller: An input device for navigation activities in virtual reality environments. In: Proceedings of the 15th ACM SIGGRAPH conference on virtual-reality continuum and its applications in industry—Volume 1. VRCAI '16. Association for Computing Machinery, New York, pp 455–460. https://doi.org/10.1145/3013971.3014005

Lindeman RW, Sibert JL, Hahn JK (1999) Hand-held windows: towards effective 2D interaction in immersive virtual environments. In: Proceedings IEEE virtual reality (Cat. No. 99CB36316). IEEE, pp 205–212

Manuri F, Piumatti G (2015) A preliminary study of a hybrid user interface for augmented reality applications. In: 2015 7th international conference on intelligent technologies for interactive entertainment (INTETAIN). IEEE, pp 37–41

Matulic F, Ganeshan A, Fujiwara H, Vogel D (2021) Phonetroller: visual representations of fingers for precise touch input with mobile phones in VR. Association for Computing Machinery, New York. https://doi.org/10.1145/3411764.3445583

Ma X, Yao Z, Wang Y, Pei W, Chen H (2018) Combining brain-computer interface and eye tracking for high-speed text entry in virtual reality. In: 23rd international conference on intelligent user interfaces. IUI '18. Association for Computing Machinery, New York, pp 263–267. https://doi.org/10.1145/3172944.3172988

Menzner T, Gesslein T, Otte A, Grubert J (2020) Above surface interaction for multiscale navigation in mobile virtual reality. In: 2020 IEEE conference on virtual reality and 3D user interfaces (VR), pp 372–381. https://doi.org/10.1109/VR46266.2020.00057

Millette A, McGuffin MJ (2016) Dualcad: integrating augmented reality with a desktop GUI and smartphone interaction. In: 2016 IEEE international symposium on mixed and augmented reality (ISMAR-Adjunct). IEEE, pp 21–26

Mine M, Yoganandan A, Coffey D (2014) Making VR work: Building a real-world immersive modeling application in the virtual world. In: Proceedings of the 2nd ACM symposium on spatial user interaction. SUI '14. Association for Computing Machinery, New York, pp 80–89. https://doi.org/10.1145/2659766.2659780

Mohr P, Tatzgern M, Langlotz T, Lang A, Schmalstieg D, Kalkofen D (2019) TrackCap: enabling smartphones for 3D interaction on mobile head-mounted displays. Association for Computing Machinery, New York, pp 1–11. https://doi.org/10.1145/3290605.3300815

Normand E, McGuffin MJ (2018) Enlarging a smartphone with AR to create a handheld VESAD (virtually extended screen-aligned display). In: 2018 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE, pp 123–133

Schmidt D, Seifert J, Rukzio E, Gellersen H (2012) A cross-device interaction style for mobiles and surfaces. In: Proceedings of the designing interactive systems conference. DIS '12. Association for Computing Machinery, New York, pp 318–327. https://doi.org/10.1145/2317956.2318005

Schwind V, Knierim P, Tasci C, Franczak P, Haas N, Henze N (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, pp 1577–1582

Sharmila A (2021) Hybrid control approaches for hands-free high level human-computer interface-a review. J Med Eng Technol 45(1):6–13

Sommer B, Wang SJ, Xu L, Chen M, Schreiber F (2015) Hybrid-dimensional visualization and interaction-integrating 2d and 3d visualization with semi-immersive navigation techniques. In: 2015 Big Data Visual Analytics (BDVA). IEEE, pp 1–8

Steed A, Julier S (2013) Design and implementation of an immersive virtual reality system based on a smartphone platform. In: 2013 IEEE symposium on 3D user interfaces (3DUI). IEEE, pp 43–46

Surale HB, Gupta A, Hancock M, Vogel D (2019) TabletInVR: Exploring the design space for using a multi-touch tablet in virtual reality. Association for Computing Machinery, New York, pp 1–13. https://doi.org/10.1145/3290605.3300243

Takashina T, Ikeya M, Tamura T, Nakazumi M, Nomura T, Kokumai Y (2018) Real-virtual bridge: Operating real smartphones from the virtual world. In: Proceedings of the 2018 ACM international conference on interactive surfaces and spaces. ISS '18, pp 449–452. Association for Computing Machinery, New York. https://doi.org/10.1145/3279778.3279921

Usoh M, Arthur K, Whitton MC, Bastos R, Steed A, Slater M, Brooks Jr FP (1999) Walking> walking-in-place> flying, in virtual environments. In: Proceedings of the 26th annual conference on computer graphics and interactive techniques, pp 359–364

Vanukuru R, Murugan A, Pillai J (2020) Dual phone AR: Exploring the use of phones as controllers for mobile augmented reality. In: 26th ACM symposium on virtual reality software and technology. VRST '20. Association for Computing Machinery, New York. https://doi.org/10.1145/3385956.3422113

Ventura J, Arth C, Reitmayr G, Schmalstieg D (2014) Global localization from monocular slam on a mobile phone. IEEE Trans Visual Comput Graph 20(4):531–539

Vinayak, Ramanujan D, Piya C, Ramani K (2016) Mobisweep: exploring spatial design ideation using a smartphone as a hand-held reference plane. In: Proceedings of the TEI '16: tenth international conference on tangible, embedded, and embodied interaction. TEI '16. Association for Computing Machinery, New York, pp 12–20. https://doi.org/10.1145/2839462.2839490

Wang S, He W, Zheng B, Feng S, Wang S, Bai X, Billinghurst M (2019) Holding virtual objects using a tablet for tangible 3D sketching in VR. In: 2019 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct). IEEE, pp 156–157

Wang S, Zhang L, Kang J, Wang S, He W (2021) Vrsmartphonesketch: Augmenting VR controller with a smartphone for mid-air sketching. In: 2021 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct), pp 362–363. https://doi.org/10.1109/ISMAR-Adjunct54149.2021.00082

Wyeth P (2008) Understanding engagement with tangible user interfaces. In: Proceedings of the 20th Australasian conference on computer-human interaction: designing for habitus and habitat, pp 331–334

Zhang L, Bai H, Billinghurst M, He W (2020) Is this my phone? operating a physical smartphone in virtual reality. In: SIGGRAPH Asia 2020 XR. SA '20. Association for Computing Machinery, New York. https://doi.org/10.1145/3415256.3421499

Zhang L, He W, Bai H, He J, Qiao Y, Billinghurst M (2021a) A hybrid 2d-3d tangible interface for virtual reality. In: ACM SIGGRAPH 2021 Posters. SIGGRAPH '21. Association for Computing Machinery, New York. https://doi.org/10.1145/3450618.3469166

Zhang L, He W, Cao Z, Wang S, Bai H, Billinghurst M (2021b) Simultaneous real walking and asymmetric input in virtual reality with a smartphone-based hybrid interface. In: 2021 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct). IEEE Computer Society, Los Alamitos, pp 321–323. https://doi.org/10.1109/ISMAR-Adjunct54149.2021.00072

Zhang L, He W, Liu Y, Zou Q, Bai H, Billinghurst M (2021c) Arinvr: Bringing mobile AR into VR. In: SIGGRAPH Asia 2021 Posters. SA '21 Posters. Association for Computing Machinery, New York. https://doi.org/10.1145/3476124.3488636

Zhu F, Grossman T (2020) BISHARE: exploring bidirectional interactions between smartphones and head-mounted augmented reality. Association for Computing Machinery, New York, pp 1–14. https://doi.org/10.1145/3313831.3376233