FocusFlow: Leveraging Focal Depth for Gaze Interaction in Virtual Reality

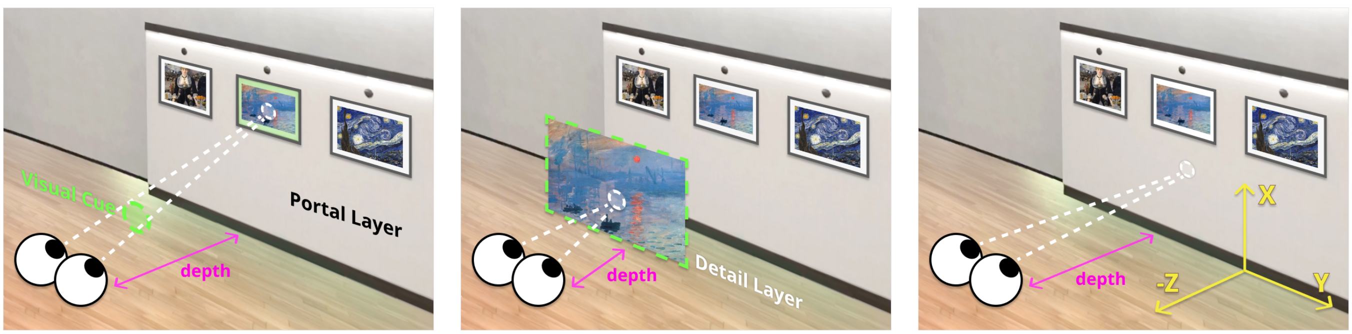
Chenyang Zhang*
zhang414@illinois.edu
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA

Tiansu Chen*
tiansuc2@illinois.edu
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA

Rohan Nedungadi
rohanrn3@illinois.edu
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA

Eric Shaffer
shaffer1@illinois.edu
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA

Elahe Soltanaghaei
elahe@illinois.edu
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA



(a) Hovering Over Target Object **(b) Activation Through Pulling In** **(c) Exit Through Pushing Out**

Figure 1: Concept of FocusFlow. (a) In our approach, we calculate the focal depth using the parallax between two eyes. When users hover over an object, a visual cue will appear to guide the users to shift their focal point between layers in different depths. (b) Users can perform hands-free layer switch to activate a panel in front by shifting their focal depth. (c) Users can exit the detail layer by pushing out the focal depth along the z-dimension.

ABSTRACT

Current gaze input methods for VR headsets predominantly utilize the gaze ray as a pointing cursor, often neglecting depth information in it. This study introduces FocusFlow, a novel gaze interaction technique that integrates focal depth into gaze input dimensions, facilitating users to actively shift their focus along the depth dimension for interaction. A detection algorithm to identify the user's focal depth is developed. Based on this, a layer-based UI is proposed, which uses focal depth changes to enable layer switch operations, offering an intuitive hands-free selection method. We also designed visual cues to guide users to adjust focal depth accurately and get familiar with the interaction process. Preliminary evaluations demonstrate the system's usability, and several potential applications are discussed. Through FocusFlow, we aim to enrich the input dimensions of gaze interaction, achieving more intuitive and efficient human-computer interactions on headset devices.

*Both authors contributed equally to this research.

UIST '23 Adjunct, October 29–November 01, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0096-5/23/10.
https://doi.org/10.1145/3586182.3615818

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; **Virtual reality**; *Interaction design theory, concepts and paradigms.*

KEYWORDS

gaze interaction, focal depth, virtual reality, user interface design

ACM Reference Format:
Chenyang Zhang, Tiansu Chen, Rohan Nedungadi, Eric Shaffer, and Elahe Soltanaghaei. 2023. FocusFlow: Leveraging Focal Depth for Gaze Interaction in Virtual Reality. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23 Adjunct), October 29–November 01, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3586182.3615818

1 INTRODUCTION

As humans, we naturally use our eyes to observe the environment and focus on objects. Taking advantage of this innate ability, a range of gaze interaction approaches have been explored for Virtual Reality (VR) and Augmented Reality (AR) headsets to enhance their input experience. Recently, Apple introduced Vision Pro, which first incorporates gaze input as a core component in commercial VR device interaction design. These gaze input methods typically utilize the gaze ray as a cursor for pointing at an object [1, 2, 8, 10, 11] and rely on other input behaviors such as hand gestures [4, 12] or dwelling [3–5] to confirm selections.

However, unlike 2D screens where the cursor provides x-y position information, our focal point in the 3D world has an additional dimension along the z-axis. Despite this depth information being available in our sight, most studies and products only consider the gaze ray direction for pointing and neglect depth information. Although some works explore the application of binocular gaze information in the VR and AR space [7, 9], none of them integrate the focal depth into current gaze interaction logic to propose a systematic interaction method. To fill this research gap, we propose FocusFlow, a novel gaze interaction method in VR that enables users to perform the layer switch by changing their focal depth with some visual guidance. Figure 1 illustrate how FocusFlow actively guides the users to shift their focal point between layers at different depths, enabling intuitive interaction with objects. This demo work contributes: (1) A sensitive and robust algorithm for detecting focal depth transitions by leveraging both eyes' gaze directions. (2) A comprehensive interaction design for "layer switch" operation. (3) Preliminary evaluations and applications demonstrating its usability and potential future usage.

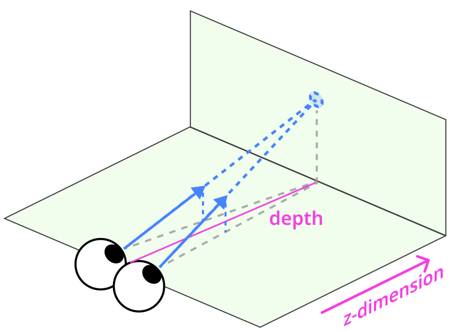


Figure 2: Depth Detection. Two gaze rays from both eyes intersect at a specific point, and the focal depth is the distance between the user and the intersection point along the z-dimension.

2 FOCUSFLOW

2.1 Detecting Focal Depth

We determine focal depth by measuring the parallax between the two eyes. To achieve this, we utilize an eye-tracker built into the VR headset to capture the positions of the eyeballs. These positions are then converted into two gaze rays originating from both eyes and intersecting at a specific point (see Figure 2). The focal depth is calculated as the distance between this intersection point and the user along the z-dimension. Additionally, since our eyes exhibit random movements when observing objects [6, 13], we have incorporated a de-noising algorithm to filter out random fluctuations in depth. For each time frame, we estimate the focal depth by current gaze directions and the depth data of recent history frames. This ensures that the resulting depth values are more stable and reliable.

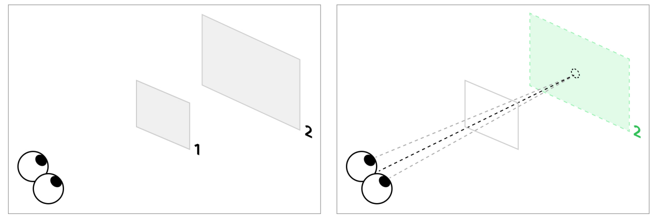


Figure 3: Layer-based UI and Depth Switch. Our proposed user interface consists of multiple layers with an information hierarchy. Users can freely switch between layers to get different levels of information by changing their focal depth.

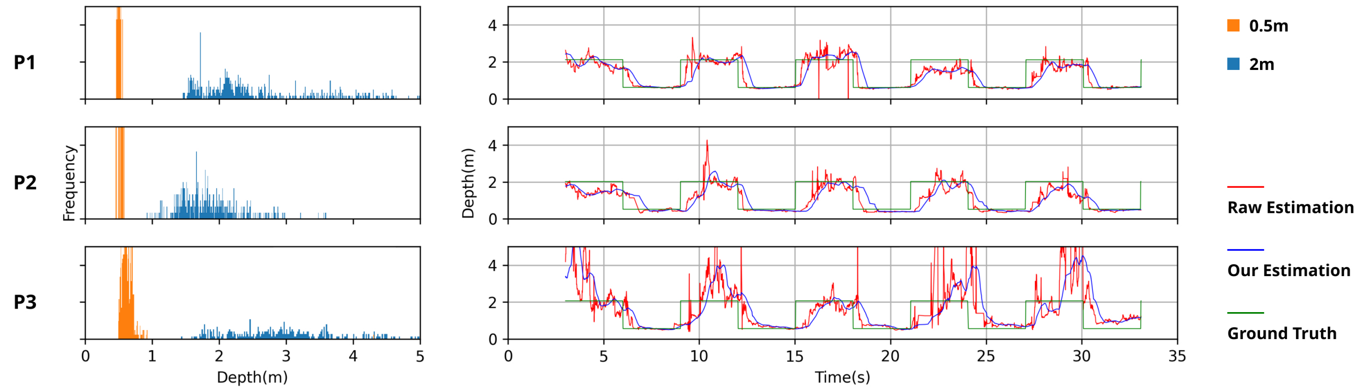
2.2 Layer Switch through Depth Change

Layer-based UI. Our focal depth changes as we observe objects at different distances. To naturally align this eye behavior with input process, we designed a layer-based User Interface (UI) that displays information on layers located at different distances along the z-dimension (Figure 3). Layers in different depths take advantage of the 3D space of the VR world and provide a new information placement space along the z-dimension. Through input in the z-dimension, users can choose any layer to view. For the sake of simplicity, we set two layers in this demo: the objects in the VR world as the portal layer (Figure 1a), and a panel located near the user as the detail layer (Figure 1b). More layers can be set and customized to more usage scenarios.

Depth Switch. Users can activate different layers by changing their focal depth. In case of our example scenario, if users want to see detailed information about an object in the portal layer, they simply need to shift their focal point towards the detail layer closer to them (Figure 1a → 1b). When users need to return to the portal layer, they can just shift their focal point further to exit the activation (Figure 1b → 1c). This provides users with an intuitive sense of "grabbing in" or "taking a closer look" at detailed information about selected targets. Similarly, shifting focal depth back towards the portal layer at the far end allows users to exit from viewing detailed information.

2.3 Visual Cue

It is worth noting that at the beginning of use, people may find it challenging to shift their focus onto something that is not visible. Therefore, it would be beneficial to offer users some cues that indicate the depth of hidden layers and assist them in adjusting the focal depth. To address this need, we incorporated a green circle in the center of the user's view as a visual cue. In the initial stages of use, users can rely on this visual cue positioned at their desired depth to quickly shift their focus and carry out interaction



(a) Depth Frequency **(b) Depth Changing Process Over Time**

Figure 4: Preliminary evaluation result from three participants. (a) Participants are asked to stare at static targets at 0.5m and 2m respectively and their focal depth data are recorded. The depth frequency shows that our depth detection algorithm can distinguish between these two different depths. (b) Participants are then asked to follow a moving target between 0.5m and 2m in distance by their eyes. The depth curve illustrates our de-noising algorithm is capable of eliminating some outliers.

operations. As users become more acquainted with the system and develop muscle memory, we gradually remove these visual cues to minimize distractions and enhance immersion.

3 USABILITY ANALYSIS

Two preliminary experiments were conducted to assess the usability and efficiency of FocusFlow. In the first experiment, participants were instructed to observe two objects in a VR scene: one located at a distance of 0.5 meters and another at a distance of 2 meters. We recorded the frequency distribution of the detected focal depth when observing the two objects as shown in Figure 4a. The results demonstrate that our detection algorithm is capable of distinguishing between two different focal depths, indicating that depth information can be collected as interaction input. However, it also reveals that for different participants, the visual depths obtained from the detection appeared to be shifted in various directions and degrees. This suggests that a personalized calibration scheme for depth detection algorithms is needed.

In the second experiment, we set up a scene where objects jump at distances of 0.5 meters and 2 meters to study the eye behavior during this transitioning process. Figure 4b illustrates how focal depth changes over time, suggesting that the eye movement is fast but with random noise present. The detected focal depth is accurate at close range, and the error becomes more significant at longer distances. Our de-noising algorithm effectively eliminates outliers and smooths the depth change curve; however, further improvement can be achieved by training additional machine learning models using eye datasets.

4 APPLICATIONS

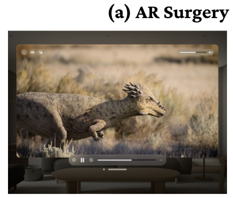
FocusFlow introduces a new dimension to interaction design, enabling the development of various applications. Here are some examples:

Hands-Free Selection. In certain scenarios, using gestures for selection may not be an optimal solution. For instance, doctors

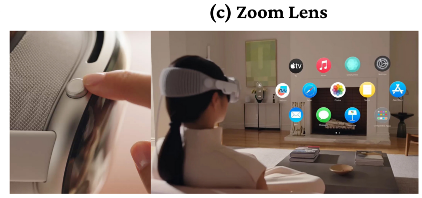
(a) AR Surgery

(b) Quick Preview

(c) Zoom Lens



(d) Video Player Bar



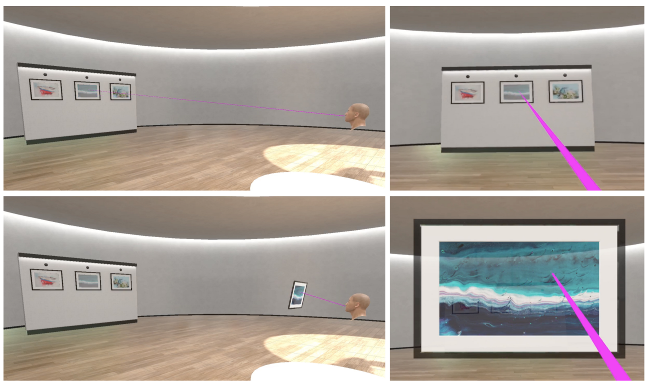
(e) Home Screen

Figure 5: Possible applications for FocusFlow.

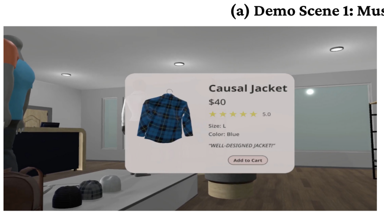
performing surgery may have their hands occupied and cannot use gestures for quick interactions (Figure 5a). Similarly, lifting one's hand for gesture recognition during simple and quick preview interactions can be cumbersome and unnecessary (Figure 5b). Moreover, individuals with disabilities may be unable to use their hands for confirmations (Figure 5c). FocusFlow offers an intuitive and efficient solution by utilizing "layer switch" as the hands-free selection method.

Activating Hidden Component. Some UI components are hidden to enhance the immersive experience in VR environments.

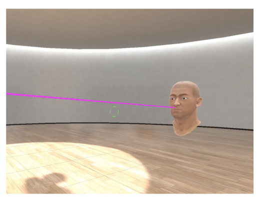
For example, when using certain applications, the home-screen is concealed, or when playing videos, the video player bar is hidden from view (Figure 5d). To activate these hidden components traditionally requires users to perform specific gestures or press designated buttons which disrupts their immersion in the virtual environment (Figure 5e). With FocusFlow, users can naturally activate these hidden components simply by shifting their attention between different depths, which is intuitive and more immersive.



(a) Demo Scene 1: Museum



(b) Demo Scene 2: Shopping



(c) Avatar in the Scene

Figure 6: Demonstration Scene in Unity.

5 DEMONSTRATION PLAN

In the demo, we will demonstrate the FocusFlow in two VR scenes, the art museum (Figure 6a) and the fashion store (Figure 6b). We adopt HTC Vive Pro Eye as our VR hardware, which enables the eye-tracking feature. Users will have an immersive experience of the FocusFlow when performing the layer switch by actively changing their focal depth.

REFERENCES

[1] Sunggeun Ahn, Stephanie Santosa, Mark Parent, Daniel Wigdor, Tovi Grossman, and Marcello Giordano. 2021. Stickypie: A gaze-based, scale-invariant marking menu optimized for ar/vr. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[2] Ramin Hedeshy, Chandan Kumar, Raphael Menges, and Steffen Staab. 2021. Hummer: Text entry by gaze and hum. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.

[3] Feiyu Lu and Doug A Bowman. 2021. Evaluating the potential of glanceable ar interfaces for authentic everyday uses. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 768–777.

[4] Feiyu Lu, Shakiba Davari, and Doug Bowman. 2021. Exploration of techniques for rapid activation of glanceable information in head-worn augmented reality. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*. 1–11.

[5] Xueshi Lu, Difeng Yu, Hai-Ning Liang, and Jorge Goncalves. 2021. itext: Hands-free text entry on an imaginary keyboard for augmented reality systems. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 815–825.

[6] Moaaz Hudhud Mughrabi, Aunnoy K Mutasim, Wolfgang Stuerzlinger, and Anil Ufuk Batmaz. 2022. My eyes hurt: Effects of jitter in 3d gaze tracking. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 310–315.

[7] Yun Suen Pai, Benjamin Outram, Noriyasu Vontin, and Kai Kunze. 2016. Transparent reality: Using eye gaze focus depth as interaction modality. In *Adjunct Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology*. 171–172.

[8] Ludwig Sidenmark, Dominic Potts, Bill Bapisch, and Hans Gellersen. 2021. Radi-Eye: Hands-free radial interfaces for 3D interaction using gaze-activated head-crossing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.

[9] Zhimin Wang, Yuxin Zhao, and Feng Lu. 2022. Control with vergence eye movement in augmented reality see-through vision. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 548–549.

[10] Xin Yi, Yiqin Lu, Ziyin Cai, Zihan Wu, Yuntao Wang, and Yuanchun Shi. 2022. Gazedock: Gaze-only menu selection in virtual reality using auto-triggering peripheral menu. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 832–842.

[11] Xin Yi, Leping Qiu, Wenjing Tang, Yehan Fan, Hewu Li, and Yuanchun Shi. 2022. DEEP: 3D Gaze Pointing in Virtual Reality Leveraging Eyelid Movement. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[12] Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2021. Gaze-supported 3d object manipulation in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[13] Xinyong Zhang. 2021. Evaluating the Effects of Saccade Types and Directions on Eye Pointing Tasks. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 1221–1234.