Inferring Gestural Intent as a Model to Support Accessible Virtual Reality Interactions for People with Upper-Limbs Motor Impairments

Siyu Zhang
s.zhang@bristol.ac.uk
University of Bristol
Bristol, United Kingdom

Kirsten Cater
kirsten.cater@bristol.ac.uk
University of Bristol
Bristol, United Kingdom

Oussama Metatla
o.metatla@bristol.ac.uk
University of Bristol
Bristol, United Kingdom

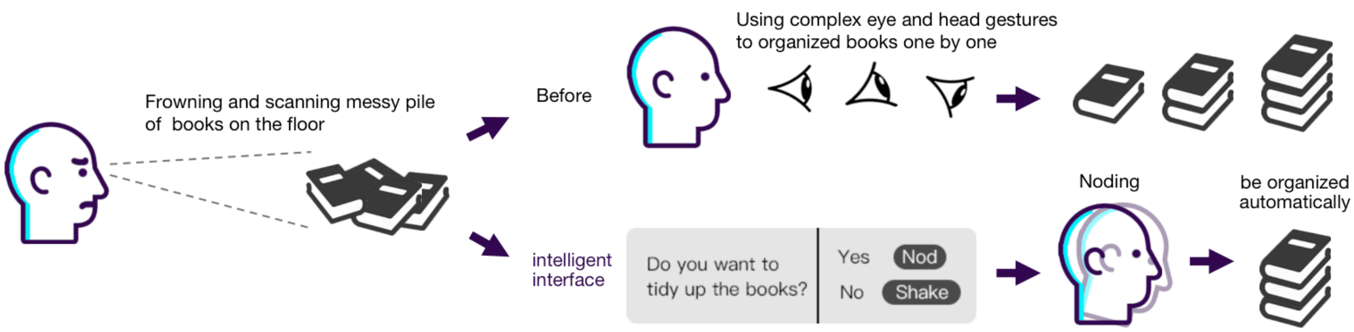

Figure 1: Example of people with upper-limb motor impairment use hand-free gestures in VR to finish complex tasks and the proposed intelligent interface to support accessibility

ABSTRACT

As Virtual Reality (VR) has the potential to help individuals with mobility restrictions to interact in ways that may not be possible in the physical world, there is an increasing demand for VR among this population. Meanwhile, people with an upper-limbs motor impairment who cannot use handheld controllers and hand gestures have numerous concerns regarding the accessibility of VR. Although there are alternative gesture-based hand-free interactions, like facial and eyelid gestures, the evaluation studies show these approaches can still not efficiently and smoothly complete complex interactions in noisy environments. This paper thus proposes to research a computational framework to predict user intent and infer users' potential intention by users' behavioral cues. Based on the estimated probabilities computed by the multivariate adaptive regression spline (MARS) model, an interaction system may display possible interactions for users to choose from. Then the system finishes user-selected interaction semi- or total- automatically. We are currently embarking on this project to explore the extent to which such an approach can simplify the interaction process in complex tasks and achieve intelligent hand-free interaction, leading to quicker, easier, more comfortable interactions for people with upper-limbs impairments.

CCS CONCEPTS

• **Human-centered computing** → **Gestural input**.

KEYWORDS

Intention recognition, Intelligent interfaces,Virtual Reality, Accessibility

ACM Reference Format:
Siyu Zhang, Kirsten Cater, and Oussama Metatla. 2018. Inferring Gestural Intent as a Model to Support Accessible Virtual Reality Interactions for People with Upper-Limbs Motor Impairments. In *Proceedings of Towards an Inclusive and Accessible Metaverse (CHI'23).* ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

CHI'23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

With the release of ever-cheaper VR systems, the number of users of virtual reality is growing. Previous studies [6] have shown that people with mobility restrictions have any number of reasons to want to experience VR. Indeed, the opportunities afforded by virtual environments are particularly beneficial to audiences who share constraints in the physical world. For example, prior work has found that people with mobility restrictions are interested in virtual experiences that might be inaccessible to them in physical world. For example, driving a buggy to explore a desert or flying gliding [5]. These transcendental experiences make someone in an accident feel much better and help them rebuild confidence after their accident [6].

VR needs to be accessible to all people, especially to users with upper-limb motor impairment. Most current interactions with VR applications receive inputs from handheld controllers or hand gestures [5], upper-limbs are essential to interact with virtual environments. For individuals with upper-limb motor impairments (either situational, temporary, or permanent), interacting with VR can be challenging, especially in manipulating dual motion controllers and inaccessible buttons [6]. Numerous researchers have investigated above-head gestures (e.g., eye gaze, head, facial, teeth, and mouth gestures) as alternative approaches to improve interactions with VR systems for people with upper-limb motor impairment. However, the evaluation results of these approaches reveal that these methods do not achieve natural low-friction interactions in many cases [8, 10]. Especially in complex tasks, users must do counter-intuitive interactions, such as concentrating for a long time to control eyelid gestures and quickly feel tired [8].

Recent researches about smart interfaces utilize statistical or machine learning models to understand users' intention. Based on the estimated probabilities computed by these models, these intelligent interaction systems can optimize the interaction of target selection [1, 4, 9], help users to avoid collisions [2] and reduce the effective latency in VR [2]. These smart interfaces with prediction methods show potential for achieving smooth and natural interactions, appearing positive effect on decreasing task completion times and error rates especially when interaction scenarios are complex and input techniques are cumbersome to use.

Inspired by this line of work, we are embarking on a project to explore an intelligent gesture interface for people who need hand-free interaction in VR. The hand-free gestures (e.g., eye gaze gestures, mouth and teeth gestures, facial gestures, head gestures, and so on) can be not only regarded as alternative input methods but also as a cue to predict users' intentions in VR. A MARS model will be built to model the relationships between users' behavior and users' intentions. We aim to build a MARS model to model the relationships between users' behavior and users' intentions, and to develop a computational framework to determine the users' intention and potential interaction. The interaction system can display recommended interactions for the user to select, then finish the selected interaction semi- or totally- automatically, allowing people with upper-limbs motor impairments to complete complex tasks with natural hand-free interaction.

2 RELATED WORKS

2.1 Accessibility of Virtual Reality for People with Upper-Limb Motor Impairments

Prior research efforts [5, 6] have evaluated present hands-free interactions and interfaces. The voice is the most studied interface, followed by the eyelid gesture and head gesture [5]. Some novel interfaces were facial gestures, mouth and teeth gestures, and brain signal interfaces. Eyelid gestures and head gestures interface show advantages on privacy and social acceptance concerns when used in public than other interface, but also have drawbacks including fatigue, low input speed and accuracy [6, 10]. In the user study of Martez Mott et al., a users is used to looking around from time to time, and the eyelid interaction in complex tasks forces them to staring at a specific area for a long time, which makes them feel unnatural [6]. Besides, another user said he need a lot of concentration to use the gaze gesture, which makes them tired so quickly that they have to exit the virtual environment early [6].Thus, it is worth investigating how best to help people with upper-body motor impairments make use of the hand-free gestures with the minimum burden of task complexity and input time of interaction.

2.2 Intention Recognition and Intelligent Suggestion in VR

Intention recognition is using behavioral cues to infer an agent's goals or future behaviour[7]. Recent research has utilized statistical or machine learning models to predict user intent and infer users' near-term interaction goals, which can drive ultralow-friction predictive interfaces.

Brendan et al. provide insight into linking human intentions and gaze cues and find gaze dynamics can predict when a user intends to interact with the digital world, which could be used to develop predictive interfaces for low-effort input[9]. Gamage et al. contribute a novel user- and activity-independent kinematics-based regressive model for continuously predicting ballistic hand movements in VR, enabling an early estimation of future events such as collisions between the user's hand and virtual objects such as UI widgets [2]. Serhan et al. used a Kalman filter to predict the viewer's head pose, allowing the estimation of future focus points of the eyes to pre-render future frames to reduce the effective latency [3]. Huang et al. developed an SVM-based approach to quantify how gaze cues may signify a person's intention, predicting which target a user might choose or interest, then use visual highlighting or display a notification to draw the user's attention towards the most probable target[4]. This method can alleviate the need to manually point at targets or conduct a full visual search of an environment, potentially leading to quicker and more comfortable interactions. Their findings offer implications for designing intention predictors for assistive systems that can provide anticipatory help to human users. However, these approaches are yet to be exploited in the design of accessible interactions in VR for people with disabilities, such as upper-limbs impairments.

3 PROPOSAL

In order to decrease accessibility barriers to hand-free gesture interaction in VR, we propose to build an intelligent gesture interface with intention recognition and intelligent suggestion. This interface regards these gestures not only the alternative input methods but also the cues to predict users' intentions and generate interaction recommendations displaying to users. The user's unconscious behavior in the virtual world becomes one of the clues to generating an instant user interface. For example, Fig. 1 shows when a user is frowning and scanning messy pile of books on the floor in a virtual world, the intelligent VR interface may infer the user's intent from the user's performance and display the interaction choices of "whether to tiding up the books". If the user nod to select "Yes", the interaction system or the user's virtual avatar will finish the action of "Tiding up the books" automatically. Such techniques could alleviate interaction tasks with complex motion curves to simple single-choice interactions, potentially leading to quicker, easier, and more comfortable interactions, and building a more accessible VR environment for people with upper-limb motor impairments.

We suggest that utilize statistical or machine learning models to understand users' intentions and generate interaction recommendations, developing and validating our prediction model through user studies. Firstly, to study how users will interact with different objects, a dense interaction tasks of different objects will be implemented in VR with 20 participants. The user behavior data will be recorded when making interaction decisions. Based on the study results, a multivariate adaptive regression splines (MARS) model will be built to model the relationships between users' behavior and users' intentions. Then a computational framework will be developed to determine the users' potential interactions, which will display in user interface.

4 GOALS OF THIS STUDY

The main goal of this study is build an accessible hand-free gesture interface with intelligent recommendation for people with upper-limb motor impairment. We suppose to achieve this goal through the following milestone goals of work:

- Modeling the relationship between users' behavior(eye gestures, facial gestures, head gestures) and users' intentions ot interact with objects by MARS.
- Building a framework to offer intelligent recommended interaction choices through a computational approach that considers users' task time and task complexity.
- Evaluating the interface with intelligent interaction recommendation by quanitifing numerically users' gesture input performance by reporting preference ratings, task completion times and error rates.

REFERENCES

[1] Brendan David-John, Candace Peacock, Ting Zhang, T Scott Murdison, Hrvoje Benko, and Tanya R Jonker. 2021. Towards gaze-based prediction of the intent to interact in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications*. 1–7.

[2] Nisal Menuka Gamage, Deepana Ishtaweera, Martin Weigel, and Anusha Withana. 2021. So predictable! Continuous 3D hand trajectory prediction in virtual reality. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 332–343.

[3] Serhan Gül, Sebastian Bosse, Dimitri Podborski, Thomas Schierl, and Cornelius Hellge. 2020. Kalman filter-based head motion prediction for cloud-based mixed reality. In *Proceedings of the 28th ACM International Conference on Multimedia*. 3632–3641.

[4] Chien-Ming Huang, Sean Andrist, Allison Sauppé, and Bilge Mutlu. 2015. Using gaze patterns to predict task intent in collaboration. *Frontiers in psychology* 6 (2015), 1049.

[5] Pedro Monteiro, Guilherme Goncalves, Hugo Coelho, Miguel Melo, and Maximino Bessa. 2021. Hands-free interaction in immersive virtual reality: A systematic review. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2702–2713.

[6] Martez Mott, John Tang, Shaun Kane, Edward Cutrell, and Meredith Ringel Morris. 2020. "i just went into it assuming that i wouldn't be able to have the full experience" understanding the accessibility of virtual reality for people with limited mobility. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–13.

[7] Ronal Singh, Tim Miller, Joshua Newn, Eduardo Velloso, Frank Vetere, and Liz Sonenberg. 2020. Combining gaze and AI planning for online human intention recognition. *Artificial Intelligence* 284 (2020), 103275.

[8] Radu-Daniel Vatavu and Ovidiu-Ciprian Ungurean. 2022. Understanding Gesture Input Articulation with Upper-Body Wearables for Users with Upper-Body Motor Impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

[9] Difeng Yu, Ruta Desai, Ting Zhang, Hrvoje Benko, Tanya R Jonker, and Aakar Gupta. 2022. Optimizing the Timing of Intelligent Suggestion in Virtual Reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–20.

[10] Xuan Zhao, Mingming Fan, and Teng Han. 2022. "I Don't Want People to Look At Me Differently" Designing User-Defined Above-the-Neck Gestures for People with Upper Body Motor Impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.