VR Book Navigation Gestures

Xiaoan Sun
The University of British Columbia
Vancouver, BC V6T 1Z4
sunxiaoan2019@163.com

Xu Fan
The University of British Columbia
Vancouver, BC V6T 1Z4
neilfanxu@gmail.com

CPSC554Y,, April 25–30, 2021, Vancouver, BC, CA
ACM xxx-x-xxxx-xxxx-x/xx/xx.
https://doi.org/xx.xxxx/xxxxxxx.xxxxxxx

Abstract
Virtual Reality (VR) is one of the most promising technologies recently. However, there is no mature study on the interaction methods of navigating through a book in VR. This study aims to design a set of hand gestures that are suitable for the book navigation problem in the Virtual Reality environment. We separated the VR book navigation task into two modes and designed a set of gestures suitable for each mode. A prototype system is implemented to study users' preferences and limitations of the designed gestures.

Author Keywords
Virtual Reality; Book Navigation Problem; Hand Gesture Design; Gesture Elicitation.

Introduction
Reading activities are one of the most common ways for people to access new information. In the real world, books are pages fastened together on one side. To navigate through books, we can turn the pages physically. But for digital libraries, we need to design an interaction method to access different contents in books. This problem is called as book navigation problem. People would develop unique book navigation interactions for each type of digital device, basing on their features. Such as PDF reader for desktops, and eReader software for tablets.

As Virtual Reality becomes more popular, the demand for its own book navigation interactions is emerging. This project is aiming to design a set of hand gestures, which are suitable for this task. By solving this problem properly, we could increase reading efficiency and the level of immersiveness in virtual reality.

In this work, we designed a set of hand gestures to fit different reading desires through the inspirations from a gesture elicitation study. A key contribution of our work is building a prototype system with part of the designed gestures. To the best of our knowledge, this is the first system that is trying to address the book navigation problem in Virtual Reality environment. Another contribution is an empirical study that explores users' preferences of book navigation hand gestures. We also highlighted the current limitations of the system and the potential future works.

To summarize, our main contributions are:

- A gesture elicitation study on the book navigation problem in VR,
- A set of hand gestures for VR book navigation problem,
- A prototype VR book navigation system,
- An empirical study highlighting the users' gesture preference.

Figure 1: Gesture for browsing mode (Left) and reading mode (Right)

Related Work

Our work relates to literature in gesture design and book navigation. This section will review the basics in these two fields.

Gesture design

Typically, creating a gesture system includes three stages [2]. First, gather requirements, ideas and decide the wanted functionalities. Second, map the functionalities to motions. Finally, perform user testing and deploy. In the middle stage, traditionally, gestures are designed by experts [12]. Another method is collecting ideas from participants through an elicitation study, which is advocated first by Wobbrock et al. [11]

To recognize gestures, there are two mainstream methods, which are vision-based, such as using Microsoft Kinect and a Leap Motion [7], and Sensor-based (Non-optical), such as using Cyber Glove [13]. Most hand model used is a replica of human's skeletal hands [10]. The hand gestures can be classified into two categories: static gestures and dynamic gestures [6]. The dynamic gestures are usually composed of three stages: start, update and end [9].

Book Navigation

To optimize the reading efficiency and user experience, researchers approached the book navigation problem in various ways. Marshall and Bly [8] performed an observational study of within-document navigation. They highlighted the absence of lightweight navigation in e-Readers. Additionally, many digital reading tools lack the function of turning multiple pages at once, which is an important part of the physical book reading experience. This work highly inspired our project.

Chen et al. [4] addressed this problem by designing a dual-display e-book reader hardware. gBook [3] projects a computer-generated book onto their hardware. Yoon et al. [14] used existing tablets and improve the interaction from the software level. They purposed Touch-Bookmark, which filled the absence of lightweight navigation in tablet E-Books. To the best of our knowledge, there is no academic approach to address the book navigation problem in Virtual Reality.

Basic Function	
1. Flip	**Flip 1 page forward/backward**
2. Checkpoint	Return to the page of last reading
3. Long Flip	**Flip multiple pages forward/backward**
4. Point	**“Point at” the content as reference**
5. Note	Make note on the content
6. Bookmark	**Mark a page for quick return**
7.Move	Adjust the position of the book
Advanced Function	
8. Zoom	**Zoom in/out the content**
9. Auto Flip	**Quick browsing the book**
10. Jump to	Go to designated page
11. Jump to directory	Go to directory page
12. Jump to chapter	Go to designated chapter
13. Select (& use add-function)	**Used for highlight, copy & paste, etc…**
Extra function	
14. Quick Flip	A simpler version of “Flip”

Table 1: 7 basic functions and 6 advanced functions in reading. Basic functions are functions concluded from either reading a physical book or an e-book. Advanced functions are functions that conclude from reading an e-book. Functions in bold type are the 7 functions selected for gesture design. “Quick Flip” is the function designed for solving gorilla arm syndrome.

Gesture Design

Function Determination

To design gestures for book navigation in VR, detailing what functions should be included is necessary. We concluded 7 basic functions and 6 advanced functions (Table. 1) that may be required in VR book reading task through analyzing what a reader may do in reading a physical book or using digital equipment (such as a tablet or personal computer) for reading. Then, we selected 7 out of 13 functions as the major objects for gesture design due to limited time and energy. The major consideration when selecting the functions is their usage frequency and basicness in the reading activity. Designing gestures for the rest 6 functions is a probable future work.

Following the gestures designing procedures mentioned by Ashbrook and Starner [2], the first stage is determining the functionality needed for the Book Navigation problem. After analyzing reading activities, we summarized 7 basic functions and 6 advanced functions, as shown in Table 1, which are potentially beneficial for the VR reading experience. Due to the time limitation, we decide to focus on the most basic functions. Then 7 functions are selected to enter the elicitation study. In the study, we would design hand gestures for each of them.

Gesture Elicitation Study

Inspired by [1], we conducted an elicitation study before designing the gesture set for functions. The aim of the elicitation study is to understand users’ preferences for a bare-handed, gesture-based book navigation system. The elicitation study is conducted on 4 master students (3M, 1F) aged 23-25.

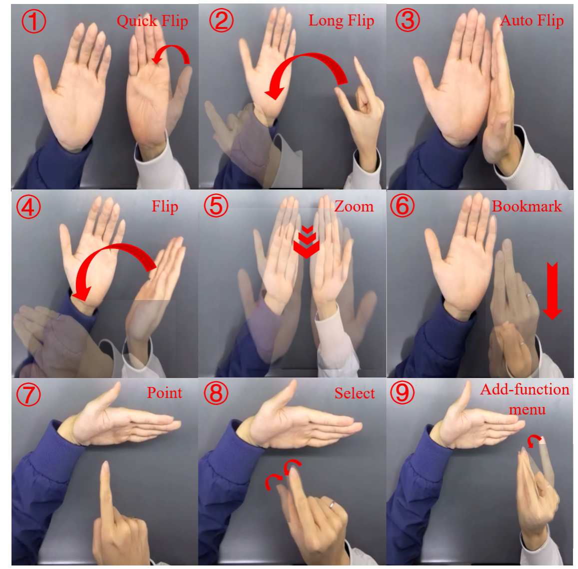

Figure 2: The hand gesture for book navigation.① ∼ ③ belong to browsing mode, ④ ∼ ⑥ belong to both modes. ⑦ ∼ ⑨ belong to reading mode. Gesture ① is a simpler version of Gesture ④ focusing on solving gorilla arm syndrome, which uses thumbs instead of a hand for flipping. Gesture ② uses the distance between the thumb fingertip and index fingertip to represent the number of pages. Gesture ③ simulates the process of pages’ natural sagging. Gesture ④ simulates the action of flipping a book. Gesture ⑤ simulates the action of moving the book toward/away from the user. Gesture ⑥ simulates the process of inserting a bookmark. Gesture ⑧ derives from ⑦ and ⑨ derives from ⑧. ⑧ and ⑨ realizes function 13 (Select (& use add-function)) together.

The elicitation study is separated into two parts. In the first part, participants listened to the description of designed functions. The description is given with only the name and

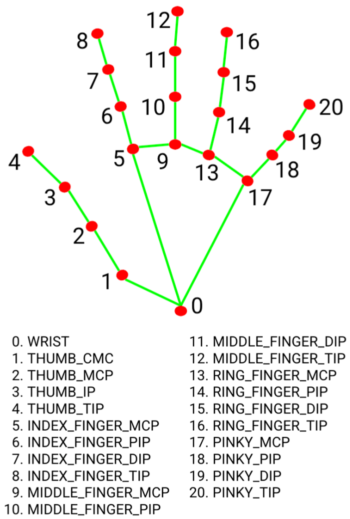

Figure 3: Hand landmarks of a hand in MediaPipe.

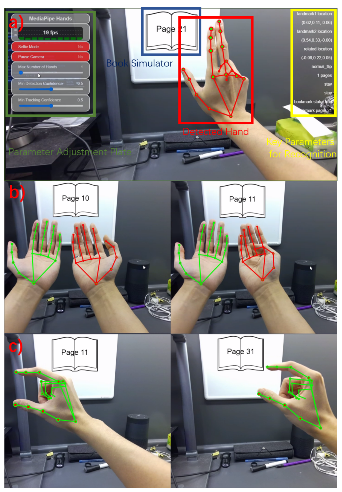

Figure 4: Book navigator system prototype. a). The entire UI of system. b). An example of using “Quick Flip”. c). An example of using “Long Flip”

brief oral explanation of the functions (without any body gesture to prevent subconscious guiding to a hand gesture) in random order. Then, they are asked to demonstrate their expected gestures for the given function. In the second part, participants are asked which functionality do they think is redundant and what other functionality are they expected to see. Sessions were video-recorded for later analysis.

Results

Based on analyzing the recorded video, we observed that participants tend to use one hand and keep the other one fixed for realizing most of the functions. For example, to flip a page, all the participants wave one of their hands and at the same time keep the other hand fixed to represent the book. Based on this finding, we decide to use single-hand movement to trigger most of our functions.

Marshall and Bly [8] found the absence of lightweight navigation in digital libraries. Inspired by this finding, we design two modes for different reading desires. Browsing mode offers the user the capability of cursory reading and reading mode allows them to do the intensive reading. The non-dominant hand will be used to switch modes. As demonstrated in Fig. 1, the user can switch between browsing mode and reading mode simply by adjusting the direction of the non-dominant hand. The dominant hand will be used to execute the functions. Each mode has three exclusive gestures and three shared gestures. Explanations can be found in Table. 1 and Fig. 2 demonstrate the entire 9 gestures. The gesture design focuses on intuitiveness, simplicity and recognizability, which is verified through a functionality mapping and memorability test. This will be explained with more details in the Experiments section.

Prototype Implementation

To test whether using hand gesture for handling book navigation problem is possible and whether our designed gestures is recognizable to computer, we developed a prototype book navigator system.

Setup

The system was implemented in JavaScript on CodePen. MediaPipe [15] is used for hand gesture recognition. A webcam is required for inputting hand gestures. No other hardware requirements are necessary.

Implementation Detail

With the help of MediaPipe, it is possible to represent a hand as a set of 21 landmarks as Fig. 3 demonstrates, which includes information such as the relative position (representing in 3 dimensions) of each landmark to the camera. Through utilizing the position of each landmark in two adjacent frames, the distance between landmarks and the movement direction of hands are calculable. Our hand gestures are recognized based on this information.

Due to the time limitation, we only implemented 4 of the most fundamental gestures of browsing mode. “Flip”, “Long Flip”, and “Bookmark” are chosen for implementation because they are the most basic functionalities for flipping through books. “Quick Flip” is implemented to increase the convenience. Fig.4 demonstrates the entire user interface and two examples about how the system works.

Experiments

The user study performed by this work consists of a functionality mapping and memorability test, a system usability test, a NASA-TLX test [5] and an interview.

Functionality mapping and memorability test
Method: 4 master students (3M, 1F) aged 23-25 were invited to the first experiment. They were asked to watch a video of a set of hand gestures and map them to the designed functionalities. We created some difficulty with the mapping test: each function may have multiple corresponding gestures; not all gestures have a corresponding function. After accomplishing the mapping test, participants received the correct answers. Five days later, they were asked to do the same mapping test again. These two tests are designed to evaluate the intuitiveness and memorability of the gestures.

Results & Discussion: Overall, the difficulty of the function mapping test is greater than our expectation. The average correctness of matching is only 62.5 percent. Participants commented, “This is a difficult task. There are more gestures than the functions in the questionnaire. Some gestures do not have a corresponding function. This increased the difficulty a lot.”. The bookmark gesture received the lowest recognition accuracy, which is 50 percent. The flip function received a low recognition accuracy as well. Most participants can correctly map the normal flip but failed in mapping the flip gesture with thumbs. Some participants reported the gesture flipping with thumbs is confusing. They prefer the normal flip gesture.

We realized that during the gesture designing process, we have already pictured the effect of using these gestures to navigate through the book. It would be more likely for us to think the gestures are intuitive. However, it would be more difficult for participants to link between functions and gestures at the first time seeing them. The results from the memorability test are significantly improved comparing with the intuitiveness test. The overall accuracy is 87.5 percent. This indicates that most of the designed gestures are easy to remember. The bookmark gesture, with a mark of 75 percent, received the lowest mark again, which means this gesture might need to be better designed. Therefore, we decide to modify the bookmark gesture design basing on users’ feedbacks.

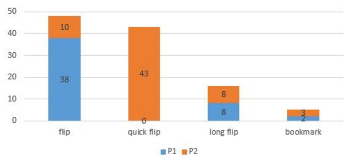

Figure 5: Gesture Usage Frequency

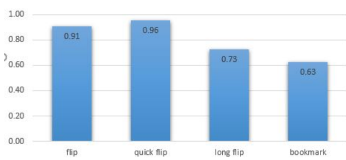

Figure 6: Recognition accuracy of the prototype system

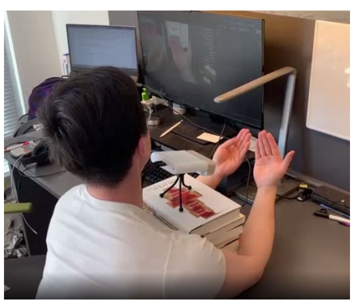
Figure 7: Experiment Setup

Usability test
Method: Two male participants are invited in this study. They first received a tutorial on how to use the system. Each of them was given 5 mins to explore it. Then they were asked to access a series of pages in sequence. The series of pages are distributed randomly. Therefore, to finish the task, participants need to navigate through the book back and forth in big movements several times. The number of gesture usage and the gesture recognition accuracy are recorded.

Results & Discussion: In the usability testing, the flip and quick flip are the most used two gestures. An interesting finding is that participant 2 prefers the quick flip gesture more than the flip. He said, “the quick flip is very convenient to flip a page. It requires much less effort than the gesture flip.” However, the other participant did not use the quick flip gesture throughout the whole test. At the end, when he was asked why he did not want to use it, he said he forgot this gesture. If he remembered, he would use the quick flip more than the flip. He commented it is quite tiring to flip many pages with the flip gesture. The quick flip would release the pain a lot.

The prototype system shows technical flaws in detection accuracy. As shown in Fig. 6, the accuracy of all the gestures needs to be further improved to provide a fluent using experience, especially for the bookmark gesture, which only has an accuracy of 66 percent. One cause of the inaccuracy could be that participants are not very familiar with the system. Sometimes their gestures are hard for the system to recognize, such as reach their hand out of the scope of the camera. Another limitation is that the vision-based recognition method is computational heavy, which causes the frame rate to decrease. Then it brings difficulty to recognize gestures accurately. The algorithm of gesture recognition can also be improved.

The conclusion drawn from this usability test is that users would prefer to choose gesture designs out of ergonomic reasons than intuitiveness. However, less intuitive designs may not come to the users in the first place. It could be helpful to remind users about the more ergonomic choice.

NASA-TLX and Interview

Method: After the prototype system usability test, participants were asked to fill in the NASA-TLX test to evaluate the workload of the task. They were also asked to share their positive and critical feedbacks to the gesture design and the prototype system.

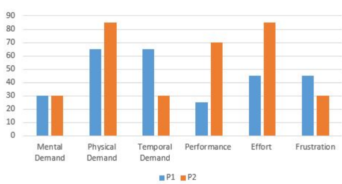

Figure 8: NASA-TLX Results

Results & Discussion: The NASA-TLX test shows on average participants think the prototype system has high physical and effort demands. They reported that it is physically tiring to follow the task. The main reason is that the experiment setup, as shown in Fig. 7, requires them to extend their arms, which is difficult to keep through the whole session. They expected that the physical demand can be significantly reduced with a more ergonomic setup. On the other hand, they reported low mental demand. They said the gesture design choices are intuitive and easy to understand.

Overall, participants showed great interest in the project. They highlighted some advantages of reading with the VR book navigator than a physical book. For example, they said "the quick flip is very convenient, even easier than flipping a real page; the long flip gives precise control on how many pages I was flipping, which I can't get from flipping with a real book.

Future Work & Limitation

Limitation

Although we have tested the system and the designed gestures on a group of participants, the range of participants is limited in both the number and the variety, which are only 4 current Asian master-level students. This may cause a bias on the reading habit and make the designed gesture not suitable for people with different educational backgrounds, ages, or cultures. Besides, our gesture and system design consider only the right-hand-based non-disabled as the target user. Therefore, how well the system and gestures will work for left-hand users or the disabled remains unknown. Moreover, since we only developed a prototype on the computer with a webcam rather than with a real VR device, whether the system can really work in VR requires to be confirmed. We also found that the frame (MediaPipe) we used in the prototype has technical flaws. For example it is sensitive to light and may mistake other objects such as the face or a textured wall as hands. This will affect the accuracy of hand gesture recognition and lead to a poor user experience.

Future Work

There are many aspects that deserve further exploration as future work.

First, it would be beneficial to extend our user study to include more participants of different races, educational backgrounds, and cultures. This will allow us to improve our gesture design to a more general version.

Second, including more functions (such as the rest 6 functions mentioned in Table. "function") for a more comprehensive gesture design and book navigation system. It would be worthy to explore combining hand gestures with traditional widget (such as push button, toggle, or slider). The balance between widget use and gesture is also a topic that deserves discussion.

Third, reducing the workload of gestures is also an important direction to go. One method is to take advantage of the fact that people would sit close to a table when reading books. Then design the gestures that would allow users to put the hands on the desk to operate.

Forth, finishing implementing all the designed gestures and build a completed gesture-based VR book navigator system. Then transfer it to Head-Mount VR devices.

Fifth, using more mature technologies to improve gesture recognition accuracy. The improvement could come from hardware, such as using Cyber Glove or Leap Motion. Or it could come from software, such as using a Machine Learning based model to recognize gestures.

REFERENCES

[1] Rahul Arora, Rubaiat Habib Kazi, Danny M. Kaufman, Wilmot Li, and Karan Singh. 2019. MagicalHands: Mid-Air Hand Gestures for Animating in VR. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 463–477. DOI: http://dx.doi.org/10.1145/3332165.3347942

[2] Daniel Ashbrook and Thad Starner. 2010. MAGIC: a motion gesture design tool. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, Atlanta, Georgia, USA, 2159. DOI: http://dx.doi.org/10.1145/1753326.1753653

[3] Jesse Burstyn and M. Anson Herriotts. 2010. gBook: an e-book reader with physical document navigation techniques. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10*. ACM Press, Atlanta, Georgia, USA, 4369. DOI: http://dx.doi.org/10.1145/1753846.1754155

[4] Nicholas Chen, Francois Guimbretiere, Morgan Dixon, Cassandra Lewis, and Maneesh Agrawala. 2008. Navigation techniques for dual-display e-book readers. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. ACM Press, Florence, Italy, 1779. DOI: http://dx.doi.org/10.1145/1357054.1357331

[5] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Eds.). Human Mental Workload, Vol. 52. North-Holland, 139–183. DOI: http://dx.doi.org/10.1016/S0166-4115(08)62386-9

[6] C. Hummels and P.J. Stappers. 1998. Meaningful gestures for human computer interaction: beyond hand postures. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*. 591–596. DOI: http://dx.doi.org/10.1109/AFGR.1998.671012

[7] Giulio Marin, Fabio Dominio, and Pietro Zanuttigh. 2016. Hand gesture recognition with jointly calibrated Leap Motion and depth sensor. *Multimedia Tools and Applications* 75, 22 (Nov. 2016), 14991–15015. DOI: http://dx.doi.org/10.1007/s11042-015-2451-6

[8] Catherine C. Marshall and Sara Bly. 2005. Turning the page on navigation. In *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries (JCDL '05)*. Association for Computing Machinery, New York, NY, USA, 225–234. DOI: http://dx.doi.org/10.1145/1065385.1065438

[9] Aditya Ramamoorthy, Namrata Vaswani, Santanu Chaudhury, and Subhashis Banerjee. 2003. Recognition of dynamic hand gestures. *Pattern Recognition* 36, 9 (Sept. 2003), 2069–2081. DOI: http://dx.doi.org/10.1016/S0031-3203(03)00042-6

[10] Vaidyanath Areyur Shanthakumar, Chao Peng, Jeffrey Hansberger, Lizhou Cao, Sarah Meacham, and Victoria Blakely. 2020. Design and evaluation of a hand gesture recognition approach for real-time interactions. *Multimedia Tools and Applications* 79, 25-26 (July 2020), 17707–17730. DOI: http://dx.doi.org/10.1007/s11042-019-08520-1

[11] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. ACM Press, Boston, MA, USA, 1083. DOI: http://dx.doi.org/10.1145/1518701.1518866

[12] Mike Wu and Ravin Balakrishnan. 2003. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *Proceedings of the 16th annual ACM symposium on User interface software and technology (UIST '03)*. Association for Computing Machinery, New York, NY, USA, 193–202. DOI: http://dx.doi.org/10.1145/964696.964718

[13] Deyou Xu. 2006. A Neural Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG. In *18th International Conference on Pattern Recognition (ICPR'06)*, Vol. 3. 519–522. DOI: http://dx.doi.org/10.1109/ICPR.2006.109 ISSN: 1051-4651.

[14] Dongwook Yoon, Yongjun Cho, Kiwon Yeom, and Ji-Hyung Park. 2011. Touch-Bookmark: a lightweight navigation and bookmarking technique for e-books. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. Association for Computing Machinery, New York, NY, USA, 1189–1194. DOI: http://dx.doi.org/10.1145/1979742.1979746

[15] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. MediaPipe Hands: On-device Real-time Hand Tracking. *arXiv:2006.10214 [cs]* (June 2020). http://arxiv.org/abs/2006.10214 arXiv: 2006.10214.