Minority Report by Lemur: Supporting Search Engine with Virtual Reality

Andrew Jie Zhou

Department of Computer Science
Georgetown University
jz398@georgetown.edu

Grace Hui Yang

Department of Computer Science
Georgetown University
huiyang@cs.georgetown.edu

ABSTRACT

In this paper, we introduce a Virtual Reality (VR) search engine interface. Virtual reality has been explored in the game industry and in the multimedia community. When wearing a VR device, a realistic experience is simulated around the user. In the working environment, VR's potential is still understudied. As a first step to enable VR-supported working environment, we present a search engine with a virtual reality interface. In our system, users can read, search and interact with the search engine with novel experiences. They only need to use their hands to interact with digital content, just like what is shown in the "minority report" movie.

KEYWORDS

Search Engine; Virtual Reality; User Interface.

ACM Reference Format:

Andrew Jie Zhou and Grace Hui Yang. 2018. Minority Report by Lemur: Supporting Search Engine with Virtual Reality. In *SIGIR '18: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, July 8-12, 2018, Ann Arbor, MI, USA*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3209978.3210179>

1 INTRODUCTION

Virtual reality (VR) has been a popular topic in our society for a while. It offers a new and exciting way of presenting digital content. Once a user is wearing a virtual reality device, a realistic experience is simulated around the user. VR's potential in technology and in marketing is far under-explored. Thanks to the availability of affordable VR devices, it is quite likely that we will witness the popularization of VR technologies [15]. In recent years, VR research has shifted its focus from primarily on hardware studies to applying the technique to various domains – i.e., a shift from hardware to applications [12]. Nowadays VR applications can be found in game [7, 16], education [1, 6, 8, 11, 13], multimedia [2, 3], and medical domains [4, 9, 10]. Unfortunately, to the best of our knowledge, little research has been done using VR in the office settings. In the 1980s, most people used papers in their offices. In the 1990s, most people used computers in their offices. We ask ourselves: Is there a possibility that people in the office would only need to use their hands to interact with digital materials, just like what is shown in

Figure 1: Curved Text Display Interface

the "minority report" movie [14]? In today's office, people spend a deal of of time using search engines. As a first step to enable a VR-supported working environment, we develop a novel VR interface for search engine. The search engine used in this research is an off-the-shelf tool, Lemur.¹ We therefore call our work "minority report by Lemur".

Regardless of VR, the design of search engine interface has not changed much in the past decade. Most search engine interfaces consist of a rectangle query box and ten blue links. Search engine algorithms heavily rely on the inputs provided by this interface. These inputs include queries wrote into the box and clicks performed over the links. They are used to determine document retrieval rankings. They are also used to generate implicit relevant judgments in search engine evaluation. We are not in a position to judge this design's effectiveness nor efficiency. However we think it might have reduced the chances to have new forms of user input signals to improve search engine performance. In this paper, we present our exploration in developing a virtual reality interface for search engine. We hope the new form of the interface would be able to introduce new forms of user input signals.

One unique experience of VR is that the users can walk around in their VR world. This would require traditional computer input accessories to move together with the users. We do not think it is desirable. Mouses and keyboards with cords are thus out of consideration. The wireless version of them might be a possibility but they would occupy the hands and limit the kinds of movements that our hands could do. For these reasons, mouses and keyboards are completely abandoned in our design. Only hands would be used to provide the inputs to the search engine.

In this research, we use a device called "Leap Motion" to capture hand movements. It contains a camera shooting at the hands and a processor analyzing the captured images. The data generated by Leap Motion are positions and rotations of hands' joints and

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SIGIR '18, July 8-12, 2018, Ann Arbor, MI, USA
© 2018 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-5657-2/18/07.
<https://doi.org/10.1145/3209978.3210179>

¹<http://www.lemurproject.org/>

bones. As an affordable device, there is definitely room to improve its detection accuracy at the hardware side. Nonetheless, given the current detection accuracy provided, we propose novel methods to optimize the gesture detection algorithm. In the end, we believe our tool is acceptable to regular search engine users.

In Hearst's book [5], she brought up a search engine interface design principle: "Search user interfaces must be understandable by and appealing to a wide variety of people of all ages, cultures and backgrounds, and for an enormous variety of information needs." We follow this principle in our design and deployment. Our contribution include the following:

- The VR interface allows a user to interact with a 3D-displayed words, queries and documents with their hands. The user would have an immense experience to touch and play with the words.
- We support new functions such as grabbing a word from the text, and merging words into new structured queries.
- We investigate ways to achieve good quality of text display with the current VR hardware.

2 DISPLAYING AND INTERACTING WITH DOCUMENTS

2.1 Displaying Text on a Curved Surface

Virtual reality is well-known for rendering impressive 3D images, however it is challenging to render text. The main reason is that displaying text requires to render too many details that are small in size. The resolution of VR devices in the market so far is too low to support high quality display of text. Besides that, text is 2D. When being shown in a 3D environment, even without any actual distortion, the user would feel the letters are stretched and distorted. A letter would also get blurry when it is displayed afar from the center view.

To find a solution to display text, we tested on several different types of VR surfaces. The choices included flat plane, sphere, and cylinder. We recruited human users to examine the visibility of displayed text. The feedback we received showed that the cylinder surface generated the best user experience.

The feedback suggested that on a flat plane, the text would be distorted at the far end of the center view. It become very difficult to be recognized visually. We excluded this design first. Both the sphere surface and the cylinder surface are curved. In general, users can see clearer text on a curved surface than on a flat one. The sphere surface and the cylinder surface differ in their effects in different view regions. For the area that is just in front of a user, i.e. the center view, the sphere surface and the cylinder surface produce similar effects. However, for the area that is above the user's view, on a sphere surface, the text would become too small to read; while on a cylinder surface, it still look normal. We therefore selected the cylinder surface to display documents.

Unlike traditional search engine interfaces, there is no top-down document list representation in our design. The top ranked document is by default displayed right in front of the user. To navigate to other documents, the user would need to use hand gestures (swiping) to reach them. In a traditional search engine interface, document positions could introduce bias because it is assumed that documents showing on the top part of the list are more relevant.

Figure 2: Swipe

Such artifacts could lead to biases that the users are more likely to click on the top positioned documents regardless their true relevance. Consequently, it might affect the effectiveness of both search engine evaluation and retrieval effectiveness. We think eliminating the top-down display order would enable us to remove this bias about ranking positions. Our new way of document display might open up opportunities for new methods of search results evaluation.

2.2 Browsing Documents by Swiping

In our system, a user can swipe their hands to navigate through the list of documents. We are aware that a user would move her hand back and forth regularly. Therefore the challenge here is to distinguish the swipe gesture used for document navigation from regular hand movements. We employ a decision tree algorithm to classify the hand movements into two classes: navigation swipe and regular hand movement. The features include movement speed, distance from the center view, distance from the last still position, and movement angle. The learned rules recognize a navigation swipe when the user's hand moves fast (passing a threshold), far from the previous still position, and perpendicular to the horizon. The classification results are used to filter out irrelevant hand movements so that a document would not be flipped unintentionally.

2.3 Grabbing a Keyword

While reading a document during search, oftentimes a user would be inspired by some words in it and would write a subsequent query. When using a desktop computer, the user could use the mouse to highlight, cutting and pasting the words into the query box to search. However, in the VR setting, we need to think new ways to allow easy cut and paste. In this particular scenario, to minimize user efforts, we enable a function to allow the user to grab a keyword to search from the current document.

The process is the following. First, we detect the event when the user's hand moves close to the curved surface that shows the document. The word which is closest to the hand would be highlighted. Next, if the user holds her hand into a fist as if the user is grabbing the word into her hand. The word that is grabbed, would create a colored (set to red) duplication of its own and move together with the fist. It is like holding the selected word in the hand. Functionally, it is equivalent to the 'copy and then drag' function on desktop. Then, the user keeps holding the fist and moves the hand towards a destination to drop the word by releasing his fist. If the destination is the place where we use to hold the queries, then the dropped word can be used as a query for next run of search.

Figure 3: Grab a Keyword

The difficulty lies in how to detect the fist opening and fist closing gestures. We investigated two approaches to detect this set of movements. The first approach is to measure and compare the distances between each hand joint. If the joints all move within a short distance to each other, then we detect the hand is making a fist. Otherwise, the hand is opening the fist. The second approach is to measure and compare the angles between the finger bones. If the angles show that the fingers are bending inwards, then we detect that the hand is making a fist. Otherwise the fist is opening. Our experiments showed that the second approach worked better. We therefore detect fist opening and closing based on bone angles.

Importantly, since a user's hands would move constantly, it is crucial to filter out movements irrelevant to fist opening and closing. Moreover, the imprecise detection made by Leap Motion would also introduce false positives when detecting positions and rotations of the bones. The challenge is to reduce those false positives. We propose to use a timeout to eliminate the false positives due to small and random hand movements. Only if a posture lasts over 0.2 seconds, we then assert it as a valid posture; and based on that we continue to determine the fist related gestures. We think other more sophisticated methods might be able to increase the gesture recognition accuracy. We leave it as future work.

3 MANIPULATING QUERIES

3.1 Inputting Keyword Queries

In our system, every input is performed by hand. Maybe in future work we could incorporate voice input. But in the current version, we must support query input only by hand. We propose a virtual keyboard to key in the words. The virtual keyboard can pop-up or hide based on hand gestures. When it is shown, the keyboard would be displayed right in front of the user on the cylinder surface. The size of the virtual keyboard is as big as half of the front view, which is easy for the user to input keyword queries.

A user can key in query keywords by tapping letters on the virtual keyboard. The tapped letters are shown in the front view too, floating in front of (closer to the user) the virtual keyboard. We could have displayed the tapped letters above or below the keyboard. However, to look at things above or below the center view, the user would have to move his head up and down drastically. Such big head movements would negatively affect the head device's detection accuracy. Therefore, we decide to show the words overlapping with the keyboard but at a different display depth. The query words are shown on a table that is below the center view.

Figure 4: Virtual Keyboard

When a user trying to tap letters to form a word, several gesture detection errors might occur and frustrate the user. Suppose the user would like to enter the letter "S". The following errors are very likely to occur. (1) She accidentally touches the left edge of letter "S". Therefore she also taps letter "A" at the same time and "A" shows up too. (2) On her way to "S", her hand accidentally touches "D" first. Therefore "D" shows up instead of "S". (3) She only wants to tap "S" once. However because of shaking of camera and/or hands, "S" appears twice. (4) She almost constantly moves her hands when reaching a sequence of different letters. For example, in the process of tapping "S" then tapping "G", she would very easily touch "D" and "F" too. Both are located between "S" and "G".

We propose two mechanisms to tackle the above problems. First, we propose to use six probes around each fingertip. Each probe is used to detect a collision with an object. Only if five or more probes are collided with the same object, we then active a tap event. Otherwise, the tap event will not be initiated. In this way, we can greatly eliminate the chance that a user taps multiple objects at the same time; the false positives are thus reduced. Second, to eliminate the error of mistyping and redundant taps, we use a cool-down timer for every object on the virtual keyboard. Each time when a key is touched, the timer starts to count down. Subtle user movements would not activate the objects unless the count down is finished.

3.2 Creating Structured Queries

Many professional users in the office settings like to use structured queries to stay control over their search results. Structured queries not only provide keywords, but also specify the relationships among the query words. Traditional search engines support structured queries using operators such as "And" and "Or". In our setting, we also support the users to create structured queries via novel VR functions. In particular, users can specify "AND" and "OR" relationships by using the following hand gestures.

3.2.1 The OR Operator. If a query has more than one word, the user can wave her hand from the top-right to the bottom-left to split the query into two parts. The resulting two parts are then in an "OR" relationship.

3.2.2 The AND Operator. A user would need to use both hands to create an AND query. The process is the following. First, she moves both of her hands next to the two queries that she would like to combine. One hand is used for one query. The two queries would turn red as the hands come close. Then, the user could grab both queries into his fists and hold them in each of the hand. The words'

Figure 5: AND query operator

backgrounds turn green to signal that they are locked. Last, the user could clap her two fists together to combine the two queries. The two keyword queries are then put into an “AND” relationship. Keywords in one AND relation can be used as a component for other “AND” or “OR” operators. We can thus form conjunctive normal form (CNF) clauses as structured search queries.

4 OTHER FEATURES

We also develop other features to aid the users while they use the new VR-enhanced search engine. Most of these features are developed for the purpose of minimizing user confusion. These features include highlighting and animation.

We add highlighting features in many parts of the VR interface to guide a user through the workflow. The consideration is that in the new VR environment, a user might feel overwhelmed by the many changes to his old experience. We therefore prompt the user with highlighting to indicate what to do next. For example, when a hand moves close to the query table, we highlight the table to indicate that the hand is very close to the table. In another example, when the hand moves inside a document, the words are highlighted one by one when they are touched.

We also use animations to reduce user confusion. A good animation can attract user attention. Without using animation, grabbing a keyword from a document would result in a ‘magical’ appearance of the grabbed word at the destination. The user might wonder “what just happened when I closed my hand?”, “why did the word suddenly appear there?” By adding animations that a grabbed word follows the hand, where the word would go becomes much more predictable and controllable. By slowing down the workflow and allowing the user to participate in the process, we manage to increase the interpretability of the system.

5 CONCLUSION & DISCUSSION

In this paper we present a novel search engine interface which is developed using virtual reality (VR). We have shown novel ways of displaying text, inputting queries and operating with structured query operators. We hope this very first academic VR-enhanced search engine would open a gate to new forms of interactions between search engine and user. We also hope VR can be used to greatly enhance searcher experience in the near future.

Most of our work focuses on supporting inputting queries and displaying the documents. We recognize these are basic functions. More advanced VR features such as showing multimedia data associated with the text would be desirable. For instance, when displaying

news articles, if we can put the user in the VR environment of the relevant videos and images, that user experience of ‘reading’ the document would be amazing. In addition, VR would create interesting and useful experience in collaborative search where multiple users could work together as easy as working face to face using the same desk. In this paper, we only focus on developing features for the front view. However, other regions of the 3D VR environment should be better utilized to support search. For instance, the users could grab and place documents with different properties at different places. Moreover, we would love to have voice input or handwriting input instead of virtual keyboard. We leave them as future work.

ACKNOWLEDGEMENTS

This research was supported by the DARPA Memex Program (grant number FA8750-14-2-0226) and NSF Career award IIS-145374. Any opinions, findings, conclusions, or recommendations expressed in this paper are of the authors, and do not necessarily reflect those of the sponsor.

REFERENCES

- [1] Eder Arroyo and Jose Luis Los Arcos. 1999. SRV: a virtual reality application to electrical substations operation training. In *Multimedia Computing and Systems, 1999. IEEE International Conference on*, Vol. 1. IEEE, 835–839.
- [2] Durand R Begault and Leonard J Trejo. 2000. 3-D sound for virtual reality and multimedia. (2000).
- [3] Chi-Cheng P Chu, Tushar H Dani, and Rajit Gadh. 1997. Multi-sensory user interface for a virtual-reality-based computeraided design system. *Computer-Aided Design* 29, 10 (1997), 709–725.
- [4] Joann Difede, Judith Cukor, Ivy Patt, Cezar Giosan, and Hunter Hoffman. 2006. The application of virtual reality to the treatment of PTSD following the WTC attack. *Annals of the New York Academy of Sciences* 1071, 1 (2006), 500–501.
- [5] Marti Hearst. 2009. *Search user interfaces*. Cambridge University Press.
- [6] Hannes Kaufmann, Dieter Schmalstieg, and Michael Wagner. 2000. Construct3D: a virtual reality application for mathematics and geometry education. *Education and information technologies* 5, 4 (2000), 263–276.
- [7] Danielle Levac, Michael R Pierrynowski, Melissa Canestraro, Lindsay Gurr, Laurean Leonard, and Christyann Neeley. 2010. Exploring children’s movement characteristics during virtual reality video game play. *Human movement science* 29, 6 (2010), 1023–1038.
- [8] Kyung-Min Park, Jeonghun Ku, Soo-Hee Choi, Hee-Jeong Jang, Ji-Yeon Park, Sun I Kim, and Jae-Jin Kim. 2011. A virtual reality application in role-plays of social skills training for schizophrenia: a randomized, controlled trial. *Psychiatry research* 189, 2 (2011), 166–172.
- [9] Sarah Parsons and Sue Cobb. 2011. State-of-the-art of virtual reality technologies for children on the autism spectrum. *European Journal of Special Needs Education* 26, 3 (2011), 355–366.
- [10] Maria T Schultheis and Albert A Rizzo. 2001. The application of virtual reality technology in rehabilitation. *Rehabilitation psychology* 46, 3 (2001), 296.
- [11] Neal E Seymour, Anthony G Gallagher, Sanziana A Roman, Michael K O’ÄZBrien, Vipin K Bansal, Dana K Andersen, and Richard M Satava. 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery* 236, 4 (2002), 458–464.
- [12] William R Sherman and Alan B Craig. 2002. *Understanding virtual reality: Interface, application, and design*. Elsevier.
- [13] KS Song and WY Lee. 2002. A virtual reality application for geometry classes. *Journal of Computer Assisted Learning* 18, 2 (2002), 149–156.
- [14] Steven Spielberg, Gerald R. Molen, Bonnie Curtis, Walter F. Parkes, Jan. Bont, Scott Frank, Jon Cohen, Tom Cruise, Colin Farrell, Samantha Morton, Max. Sydow, Lois Smith, Peter Stormare, Tim B. Nelson, Steve Harris, Kathryn Morris, and Philip K. Dick. 2003. *Minority report*. (2003).
- [15] Telmo Zarraonandia, Paloma Díaz, Ignacio Aedo, and Alvaro Montero. 2016. Immersive End User Development for Virtual Reality. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 346–347.
- [16] Michael Zyda. 2005. From visual simulation to virtual reality to games. *Computer* 38, 9 (2005), 25–32.