GazeGrip: Improving Mobile Device Accessibility with Gaze & Grip Interaction

Qiushi Zhou
Microsoft Research Centre for Social NUI
University of Melbourne
haust798@gmail.com

Eduardo Velloso
Microsoft Research Centre for Social NUI
University of Melbourne
eduardo.velloso@unimelb.edu.au

[1]ABSTRACT

Though modern tablet devices offer users high processing power in a compact form factor, interaction while holding them still presents problems, forcing the user to alternate the dominant hand between holding and touching the screen. In this paper, we explore how eye tracking can minimize this problem through GazeGrip---a prototype interactive system for a tablet that integrates eye tracking and back-of-device touch sensing. We propose a design space for potential interaction techniques that leverage the power of this combination, as well as prototype applications that instantiate it. Our preliminary results highlight as opportunities enabled by the system reduced fatigue while holding the device, minimal occlusion of the screen, and improved accuracy and precision in the interaction.

CCS CONCEPTS

• **Human-centered computing → Interaction Techniques**

KEYWORDS

Gaze input; Back-of-device Interaction; Multimodal UI; Tablet.

ACM Reference format:

Q. Zhou and E. Velloso. 2017. GazeGrip: Improving Mobile Device Accessibility with Gaze & Grip Interaction. In *Proceedings of the 29th Australian Conference on Human-Computer Interaction, Brisbane, QLD, Australia, November 2017 (OzCHI 2017)*, 4 pages.
https://doi.org/10.1145/3152771.3156159

1 INTRODUCTION

Recent advances in technological miniaturization and battery life have turned tablet devices into general purpose computers. Though tablet devices can be used in the same way as laptops on a desk, the interaction is still awkward when users are holding the device. Due to their size and weight, people use different postures to support and to interact with them, with a trade-off between comfort and degrees of freedom for interaction [14]. For most interaction tasks with a tablet, users must constantly switch the role of the dominant hand between *holding* the device and *touching* the screen. In this paper, we present a system that aims at minimizing this problem through a combination of gaze and grip interaction (see Figure 1). **GazeGrip** is a prototype interactive system built around the Microsoft Surface Pro 3 that integrates eye tracking with back-of-device touch sensing, enabling users to interact with the device while holding it with both hands, hence minimizing the fatigue of holding it with one hand, the restriction for interaction while holding it with both hands and the occlusion caused when touching the screen.

Figure 1: Interacting with GazeGrip: With Gaze and Back-of-Device touch sensing, users are able to interact with the tablet while holding it with both hands without occlusion.

To explore possible interaction techniques afforded by this combination, we propose a design space in which each modality can act in a *passive* or *active* manner. We instantiate this design space with three applications: a text reader, a digital camera and a text entry system. These applications highlight several opportunities for the combination of gaze and grip in large tablet devices, including minimizing the fatigue of holding the device, enabling users to always hold the device with both hands, eliminating the occlusion caused by touching the screen, and enhancing the accuracy of the interaction.

In summary, the contributions of this paper are as follows: (1) a multimodal interactive system, realized by the design and implementation of a novel case for large tablets that integrates eye tracking and back-of-device touch sensing; (2) a design space of potential interaction techniques for the system; (3) three applications that demonstrate the opportunities for improving the interaction with tablet devices; (4) a heuristic evaluation.

2 RELATED WORK

Gaze presents a wealth of opportunities as an input modality for interactive systems while facing several challenges, as previous works highlighted. Gaze is often combined with other modalities

OzCHI '17, November 28-December 1, 2017, Brisbane, QLD, Australia

ACM 978-1-4503-5379-3/17/11..$15.00
https://doi.org/10.1145/3152771.3156159

due to the Midas Touch problem [5], such as the mouse [20] or keyboard [6]. With the popularity of touch-sensitive screens, researchers started to explore how gaze can be incorporated into the interaction with mobile devices. Stellmach et al. combined gaze with touch for target acquisition on multiple devices, following the principle *gaze suggests, touch confirms* [12]. Pfeuffer et al. proposed interaction techniques for combining gaze and touch on the same surface, enabling seamless switch between direct and indirect manipulation [8]. While these works have explored the combination of gaze with touch on screens, in this paper we explore its combination with simple gestures performed by the gripping fingers on the back of mobile devices.

Back-of-device interaction has mostly been studied as a complement for front-of-device interaction. Sugimoto and Hiroki designed an interaction technique for mobile devices called *HybridTouch* showing that users are able to perform interactive tasks intuitively while operating on the front and back of the device at the same time. A taxonomy for *microinteractions* proposed by Wolf et al. suggested that back-of-device interaction in the form of gripping gestures is ergonomically feasible [19]. Wolf et al. later proposed *PinchPad* as two multitouch tablets put together back-to-back to simulate a system with both front and back touchable surfaces. It suggested that gestures comprised of finger movements are recognizable on the back of the device, allowing more stability while holding the device with both hands [18]. Compared to front-of-device interaction, back-of-device interaction allows for natural hand posture while suffering from poor perception [1] and occlusion [15]. In this paper, we explore how gaze can be used to compensate for this problem.

Napier proposed a taxonomy of human grip gestures and concluded that humans are able to perform *precision grips* intuitively to accomplish complicated tasks [7]. Saponas et al. suggested that **hand-busy interaction** is possible when the hands of users are already engaged in a task or supporting a heavy load [9]. Wimmer and Boring presented a prototype called *HandSense* to demonstrate that grip information is able to enhance implicit and explicit interaction [17]. Schmidt defined **implicit interaction** as an action performed by the user that is not primarily aimed to interact with a computerized system but which such a system understands as input [10]. Hansen et al. categorized the applications of gaze control into active device control and passive recording [4]. Inspired by this categorization, in the context of holding tablet device with both hands, we define *active interaction* as an instance in which users actively express their interests to the interactive system. Conversely, we define *passive interaction* as the monitoring of users' affective reactions triggered by the system, indicating their spontaneous emotional reactions and mental involvement. In GazeGrip, we first explore the integration of gaze and grip acting as active and passive interaction to provide interactive context for each other with the aim of improving accessibility for tablet (see Figure 1).

3 DESIGN SPACE

To explore the possible interaction techniques afforded by the combination of gaze and grip interaction, we devised a design space in which each of them acts either as a passive or as an active input modality (see Table 1). This is motivated by the fact that both how we look at the screen and how we hold the device are informative of the context surrounding the user and the activities they are engaged in. In this section, we explore each cell of the design space in general and suggest potential applications for them.

Table 1: A Design Space for GazeGrip

	Passive Grip	Active Grip
Passive Gaze	**Context Awareness** Both Gaze and Grip as passive modalities to interpret context	**Gaze-Contingent Grip** Gaze provides context Grip triggers operation
Active Gaze	**Grip-Contingent Gaze** Grip provides context Gaze triggers operation	**Gaze selects, Grip confirms** Both Gaze and Grip accept instructions from user

3.1 Context Awareness

This condition employs both gaze and grip as passive modalities, monitoring both the manual (how the user is holding the device) and visual (user's gaze pattern) behaviour to extract context information. *Context awareness* enables attention detection while holding tablet with both hands. In the context of this work, gaze and grip are the only two input modalities utilized. Thus, users are assumed to have little or no attentive involvement with the system when no active or reactive interaction is detected over time in either of the two modalities.

3.2 Grip-Contingent Gaze

In this condition, passive grip provides context for active gaze. This cell comprises applications where the system is able to provide different graphical interfaces to users when their grip gestures change unconsciously. Users are then able to use gaze to interact with the interface. For example, scrolling is crucial in applications using non-fixed screen space. In the context of holding tablet with both hands, the user's active gaze is able to control the direction and speed of scrolling without touching on the surface of the tablet, while passive grip provides context for the system to determine when the scrolling function is intended.

3.3 Gaze-Contingent Grip

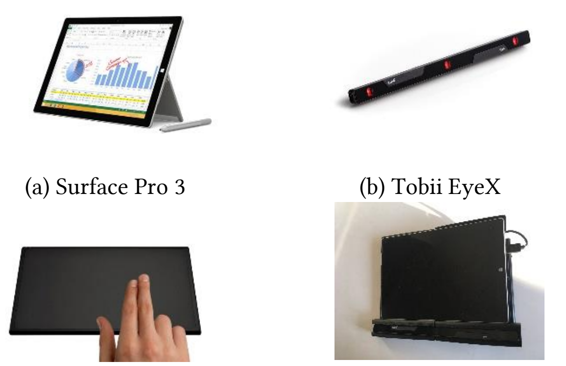
Passive gaze provides context for active grip to achieve content-aware interaction. This combination is suitable for simple tasks in which users instruct the system to perform operations on the area of interest. This combination inspires a photo capturing function in which the extra step of selecting the focal point is replaced by discovering the point of interest, hence reduces physical effort, minimizes visual occlusion and avoids the Fat Finger problem [11].

3.4 Gaze selects, Grip confirms

Active gaze input and active grip provide maximum control to users in the context of holding tablet with both hands. This combination helps users complete more complex tasks and works in a similar way with *Look & Touch* following the principle gaze suggests, grip confirms [12]. It enables typing function with both hands holding a tablet. It simplifies hand-eye coordination in traditional typing functions on touch screens by reducing the touching step to a gripping gesture. Users are able to type by gripping the device while looking at the intended letter. This way of text input improves accessibility and efficiency by utilizing the busy hand. It minimizes physical effort by reducing hand movement required by traditional typing.

4 HARDWARE DESIGN

The physical prototype of GazeGrip comprises of a Microsoft Surface Pro 3, a Tobii EyeX eye tracker, a Kogan Magic Trackpad, a USB Hub and a 3D-printed case binding all the components (Figure 2). The 3D model of the case is presented in Figure 3. We designed the case with the aim of increasing the ease of holding it with both hands. The dimensions of the case are 344 x 242 x 46 mm. The front of the case comprises an upper part accommodating the Microsoft Surface Pro 3, and a lower part to allow the Tobii EyeX to slide in from the side.



(a) Surface Pro 3 (b) Tobii EyeX

(c) Kogan Trackpad (d) GazeGrip with Case

Figure 2: Hardware Components of GazeGrip

(a) Side view (b) Back view

Figure 3: 3D Model of the Case

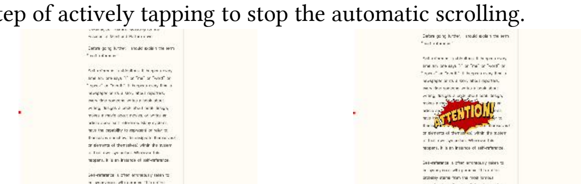
The Kogan Magic Trackpad slides from the side into an extra 126 x 203 mm slot attached on the back of the case. A 71 x 87 mm space was cut on the back of the slot to allow different grip gestures in forms of touching and tapping gestures on the trackpad(Figure 3(b)). We reserved a 25 mm-wide edge next to the touching space with the aim of minimizing false touch interaction(Figure 3(b)). We added an extra slot on the back to accommodate the Targus 4-port USB Hub to connect the devices (Figure 1).

5 APPLICATIONS

We present design and development details of three applications, each instantiating one or more functions suggested in the design space. Screenshots of 3 applications are illustrated in Figure 4–6.

5.1 Application: GazeGrip Reader

The GUI of GazeGrip Reader presents users with the content at the center of the screen (Figure 4). In GazeGrip Reader, users start reading by double tapping on the trackpad without lifting the fingers to trigger the automatic scrolling function. The reading material moves up when users' gaze positions are below the center of the screen, and vice versa (Figure 4(a)). The scrolling accelerates when user looks towards the horizontal edges of the screen, and it stops when users' gripping fingers leave the trackpad. GazeGrip Reader enables users to quickly respond to interruptions during reading tasks without the extra step of actively tapping to stop the automatic scrolling.



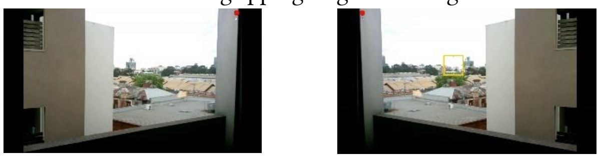
(a) Automatic Scrolling (b) Attention Alert

Figure 4: GazeGrip Reader

GazeGrip Reader instantiates *Grip-Contingent Gaze* to enable gaze interaction with the context information provided by passive grip. The scrolling automatically pauses as the system interprets users' passive intention of suspending the reading task by recognizing users' passive touch behaviour on the trackpad. GazeGrip Reader also instantiates *Context Awareness*. The system initiates an attention alert when it detects no horizontal movement of the user's gaze on the screen over 5 seconds by displaying an alert icon (Figure 4(b)), in the condition that the gripping fingers don't move. GazeGrip Reader utilizes both eye tracking data and back-of-device touching data as contextual information to implement attention detection.

5.2 Application: GazeGrip Camera

The GUI of GazeGrip Camera is comprised of a camera view taking up the entire screen (Figure 5(a)). GazeGrip Camera determines the focal point of a picture automatically for users by interpreting their gaze patterns. The system chooses a location as the point of interest when users' gaze positions stay within 100 pixels over 1 second. An animation of a scaling frame informs the user where the focal point has been chosen (Figure 5(b)). Users are able to take a picture by tapping on the back of the case with one of the gripping fingers holding the device.



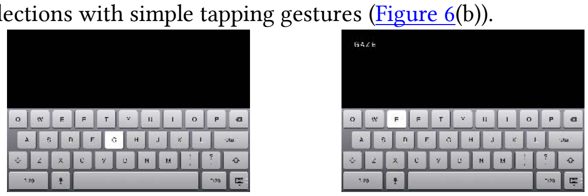
(a) Camera View (b) Focusing Animation

Figure 5: GazeGrip Camera

GazeGrip Camera instantiates *Gaze-Contingent Grip* by interpreting users' passive gaze patterns to reduce physical effort required in traditional photo capturing tasks where users have to manually choose the focal point with one hand while holding the device with the other hand.

5.3 Application: GazeGrip Type

The GUI of GazeGrip Type is comprised of a virtual keyboard at the bottom of the screen with a text displaying area on top (Figure 6). Intended letters are highlighted when users' gaze positions are within the boundaries of the keys (Figure 6(a)). Users are able to confirm the selection by tapping on the trackpad with one of the gripping fingers. Users are able to type by scanning through the intended letters while confirming the selections with simple tapping gestures (Figure 6(b)).



(a) Highlight Selection (b) Continuous Typing

Figure 6: GazeGrip Type

GazeGrip Type instantiates *Gaze selects, Grip confirms*. It utilizes both gaze and grip in active forms to reduce physical effort of touching with one hand while holding the device with the other hand. GazeGrip Type also minimizes visual occlusion caused by the touching hand.

6 DISCUSSION

In this section, we discuss the benefits of GazeGrip considering physical aspects, occlusion and accuracy of interaction. As the system is still in an early prototypical stage, a heuristic evaluation such as this is more appropriate for assessing future opportunities of usage than an empirical evaluation.

6.1 Physical aspects

GazeGrip minimizes the physical fatigue induced by supporting the device with one hand by enabling interaction with both hands holding the device. GazeGrip also increases the stability of the device by enabling users to interact with the device while holding it with both hands.

GazeGrip saves users from fatigue [2] and Musculoskeletal Disorders [3] caused by traditional interaction techniques with touch screens. GazeGrip Reader reduces the fatigue caused by excessive vertical movement of the interacting hand in manual scrolling tasks. GazeGrip Type allows users to type with less physical effort by enabling them to type without moving their hands over the entire virtual keyboard.

6.2 Occlusion

GazeGrip minimizes visual occlusion by moving the touch interaction to the back of the device. This shift provides users with visual access to the entire screen. GazeGrip Camera reduces the occlusion caused by manually selecting the focal point on the screen. Gaze Reader reduces occlusion by enabling users to scroll without their hands over the screen. Users are allowed to have visual access to any content presented on the screen at any time.

6.3 Accuracy of interaction

Compared with gaze-only input, GazeGrip allows more accurate interaction. GazeGrip simultaneously minimizes the Midas Touch problem and the Fat Finger problem, as touch confirms gaze actions, and by redirecting touch events to the back of the device. A combination of gaze pointing and touch-based fine tuning enables precise selection of small targets. Grip improves accuracy of interaction by providing a confirming modality that is simple enough to allow minimal probability of error, while gaze utilizes eye tracking technology with high-resolution. GazeGrip integrates gaze and grip to enable interaction that is both accurate and precise.

7 LIMITATIONS

We introduced GazeGrip as a novel interactive system for the specific context of holding tablet with both hands. In this work, we focused on analyzing and demonstrating the design space instead of conducting an empirical evaluation at this stage, as inspired by previous works with the same approach [16, 13].

8 CONCLUSION

We proposed GazeGrip as a novel system that affords the combination of gaze and grip interaction in active and passive ways. Based on a heuristic assessment, we noted that GazeGrip was comfortable to use, offers a solution to the problem of occlusion in back-of-device touch interaction, and enhances stability by allowing users to interact with tablets while gripping the device with both hands. We proposed a design space that inspired the design of three applications. Future works can incorporate more modalities such as accelerometers, light sensors, etc. to extend the feature set available for the design space.

REFERENCES

[1] Patrick Baudisch and Gerry Chu. 2009. Back-of-Device Interaction Allows Creating Very Small Touch Devices. In *Proceedings of Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 1923-1932. DOI: http://dx.doi.org/10.1145/1518701.1518995

[2] Radhika Bhatt, Nikita Fernandes and Archana Dhage. 2013. Vision Based Hand Gesture Recognition for Human Computer Interaction. *International Journal of Engineering Science and Innovative Technology (IJESIT)* 2, 3, 110-115. Retrieved from: http://www.ijesit.com/Volume%202/Issue%203/IJESIT201303_17.pdf

[3] JE Gold, JB Driban, Nadja Thomas, Tapash Chakravarty, V Channell and Eu-gene Komaroff. 2012. Postures, Typing Strategies, and Gender Differences in Mobile Device Usage: An Observational Study. *Applied ergonomics* 43, 2, 408-412. DOI: https://doi.org/10.1016/j.apergo.2011.06.015

[4] John Paulin Hansen, Allan W Andersen and Peter Roed. 1995. Eye-Gaze Control of Multimedia Systems. *Advances in Human Factors/Ergonomics* 20, 37-42. DOI: https://doi.org/10.1016/S0921-2647(06)80008-0

[5] Robert JK Jacob. 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at Is What You Get. *ACM Transactions on Information Systems (TOIS)* 9, 2, 152-169. DOI: https://doi.org/10.1145/123078.128728

[6] Manu Kumar, Andreas Paepcke and Terry Winograd. 2007. Eyepoint: Practical Pointing and Selection Using Gaze and Keyboard. In *Proceedings of Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM, 421-430. DOI: https://doi.org/10.1145/1240624.1240692

[7] John R Napier. 1956. The Prehensile Movements of the Human Hand. *Bone & Joint Journal* 38, 4, 902-913. DOI: https://doi.org/10.1007/978-1-4471-5451-8_85

[8] Ken Pfeuffer, Jason Alexander, Ming Ki Chong and Hans Gellersen. 2014. Gaze-Touch: Combining Gaze with Multi-Touch for Interaction on the Same Surface. In *Proceedings of Proceedings of the 27th annual ACM symposium on User interface software and technology.* ACM, 509-518. DOI: https://doi.org/10.1145/2642918.2647397

[9] T Scott Saponas, Desney S Tan, Dan Morris, Ravin Balakrishnan, Jim Turner and James A Landay. 2009. Enabling Always-Available Input with Muscle-Computer Interfaces. In *Proceedings of Proceedings of the 22nd annual ACM symposium on User interface software and technology.* ACM, 167-176. DOI: https://doi.org/10.1145/1622176.1622208

[10] Albrecht Schmidt. 2000. Implicit Human Computer Interaction through Context. *Personal technologies* 4, 2, 191-199. DOI: https://doi.org/10.1007/BF01324126

[11] Katie A Siek, Yvonne Rogers and Kay H Connelly. 2005. Fat Finger Worries: How Older and Younger Users Physically Interact with Pdas. In *Proceedings of INTERACT.* Springer, 267-280. DOI: https://doi.org/10.1007/11555261_24

[12] Sophie Stellmach and Raimund Dachselt. 2012. Look & Touch: Gaze-Supported Target Acquisition. In *Proceedings of Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2981-2990. DOI: https://doi.org/10.1145/2207676.2208709

[13] Eduardo Velloso, Carl Oechsner, Katharina Sachmann, Markus Wirth and Hans Gellersen. 2015. Arcade+: A Platform for Public Deployment and Evaluation of Multi-Modal Games. In *Proceedings of Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play.* ACM, 271-275. DOI: https://doi.org/10.1145/2793107.2793145

[14] Julie Wagner, Stéphane Huot and Wendy Mackay. 2012. Bitouch and Bipad: Designing Bimanual Interaction for Hand-Held Tablets. In *Proceedings of Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2317-2326. DOI: https://doi.org/10.1145/2207676.2208391

[15] Daniel Wigdor, Darren Leigh, Clifton Forlines, Samuel Shipman, John Barnwell, Ravin Balakrishnan and Chia Shen. 2006. Under the Table Interaction. In *Proceedings of Proceedings of the 19th annual ACM symposium on User interface software and technology.* ACM, 259-268. DOI: https://doi.org/10.1145/1166253.1166294

[16] Andrew D Wilson. 2004. Touchlight: An Imaging Touch Screen and Display for Gesture-Based Interaction. In *Proceedings of Proceedings of the 6th international conference on Multimodal interfaces.* ACM, 69-76. DOI: https://doi.org/10.1145/1027933.1027946

[17] Raphael Wimmer and Sebastian Boring. 2009. Handsense: Discriminating Different Ways of Grasping and Holding a Tangible User Interface. In *Proceedings of Proceedings of the 3rd International Conference on Tangible and Embedded Interaction.* ACM, 359-362. DOI: https://doi.org/10.1145/1517664.1517736

[18] Katrin Wolf, Christian Müller-Tomfelde, Kelvin Cheng and Ina Wechsung. 2012. Pinchpad: Performance of Touch-Based Gestures While Grasping Devices. In *Proceedings of Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction.* ACM, 103-110. DOI: https://doi.org/10.1145/2148131.2148155

[19] Katrin Wolf, Anja Naumann, Michael Rohs and Jörg Müller. 2011. A Taxonomy of Microinteractions: Defining Microgestures Based on Ergonomic and Scenario-Dependent Requirements. In *Proceedings of IFIP Conference on Human-Computer Interaction.* Springer, 559-575. DOI: https://doi.org/10.1007/978-3-642-23774-4_45

[20] Shumin Zhai, Carlos Morimoto and Steven Ihde. 1999. Manual and Gaze Input Cascaded (Magic) Pointing. In *Proceedings of Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* ACM, 246-253. DOI: https://doi.org/10.1145/302979.303053