Combining Touch and Gaze for Distant Selection in a Tabletop Setting

Michael Mauderer
School of Computer Science,
University of St Andrews
mm285@st-andrews.ac.uk

Antonio Krüger
German Research Center for
Artificial Intelligence (DFKI)
krueger@dfki.de

Florian Daiber
German Research Center for
Artificial Intelligence (DFKI)
florian.daiber@dfki.de

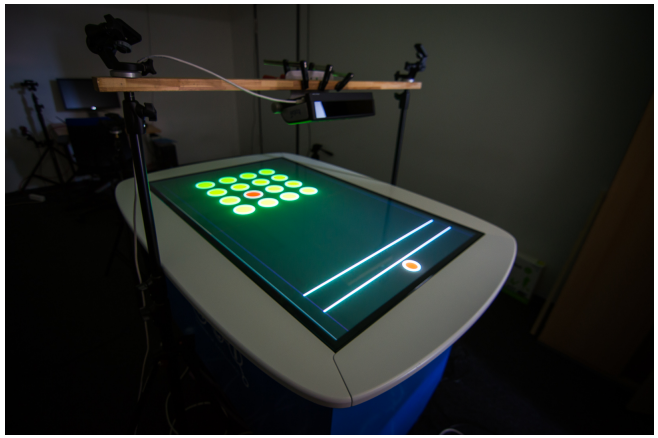

Figure 1: Overview of the used setup.

CHI 2013 Workshop on "Gaze Interaction in the Post-WIMP World", April 27, 2013, Paris, France.

Abstract
Tabletop interaction with objects in and out of reach is a common real world as well as virtual task. Gaze as additional input modality might support this interactions on tabletops in terms of search, selection and manipulation of distant objects. The aim of this work is to design and evaluate an interaction technique that relies on gaze and gestural touch input for the selection of distant objects. The proposed approach makes objects that are out of physical reach easily available to the user, and aims to provide an increased selection accuracy compared to single modality approaches. The paper contributes a setup that allows to track people with a static eye-tracker in front of a tabletop and investigates an interaction technique that makes use of the flicking gesture augmented by gaze information to select distant objects.

Author Keywords
Gaze input; touch interaction; selection; flicking; gaze-supported interaction

ACM Classification Keywords
H.5.2 [Information interfaces and presentation]: User Interfaces. Input devices and strategies

General Terms
Design, Human Factors

Introduction and Motivation

Multi-touch enabled tabletops are becoming increasingly popular as interactive information displays. Especially with larger tabletops there often is the need to interact with objects that are out of reach. In the past there have been different approaches proposed to solve this problem; some of them relying on gaze tracking. But using gaze alone has been shown to be problematic for this task due to the well known *Midas Touch* effect [4]. This led to the use of additional input modalities to support gaze-based interactions. Following this approach, the aim of this work is the design and evaluation of an interaction technique that relies on gaze and gestural touch input. In particular the selection of distant objects on a large tabletop will be investigated. Since tabletops are often located in public spaces and meant to be easily accessible, eye tracking in this context should also be as unobtrusive as possible. Therefore a static eye tracker will be used; it does not need instrumentation of the user and it can be integrated in the environment. The proposed approach aims to combine the natural interaction of a simple gesture with the increased accuracy of additional gaze data.

Based on the previous considerations the contribution of this work is two-folded: First, we propose a solution for an interactive eye tracking-enabled tabletop setup usable for interaction studies. Secondly, an interaction technique for distant selection, which combines gaze and flicking gesture, is proposed and evaluated in an user study.

The remainder of this paper is organized as follows: First related approaches from gesture and gaze interaction are explored. Then the concept of gaze and flick selection and a prototype is presented. Afterwards, a study to evaluate the system is presented and the results are discussed. Finally, a conclusion and outlook on future work is given.

Related Work

For interaction with distant targets on tabletops a number of purely (multi-) touch based interaction techniques have been proposed. The *I-Grabber* [1] uses the metaphor of a physical grabber to reach for objects. *Gesture Select* [3] combines an pointing gesture with drawing of markers for selection of annotated objects. *Superflick* [5] allows to slide virtual objects on the tabletop surface towards their target using a flicking gesture.

When using eye tracking for interaction the gaze position is most commonly used as a simple cursor, as already done by Ware [9], who proposed dwell time based approaches, as well as manual selection of attended targets. Since gaze-only selection through dwell time is prone to the Midas Touch effect [4] alternative approaches have been investigated. Among them Bader *et al.* [2] who proposed the integration of natural gaze behaviour in different contexts to determine the user's intentions during object manipulation. Other multi-modal approaches try to use gaze in supporting roles in combination with other selection techniques. *Magic Pointing* [10] uses the mouse as primary selection method, but tries to increase selection speed by displacing the cursor position towards the attended target. While the aforementioned approaches were designed to be used in a desktop setting *Gaze Galaxy* [7] and *Look and Touch* [6] uses gaze-supported interaction on distant displays. The main interaction here is done through additional hardware, i.e. mobile (touch) devices or a keyboard, and several selection techniques featuring different uses of touch and gaze are presented. Turner *et al.* [8] propose several interaction techniques for distant interaction on large multi-touch surfaces and public displays using active gaze interaction in combination with several touch gestures, which are executed either on the surface itself or on a mobile device.

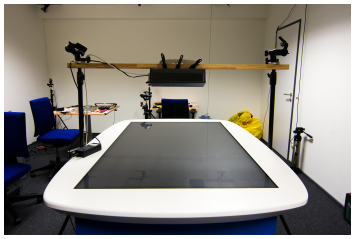

Figure 2: Tabletop setup with mounted eye tracker; seen from the front.

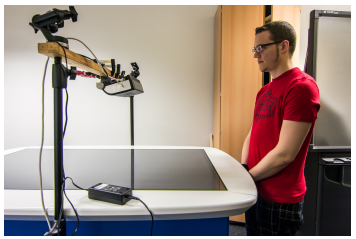

Figure 3: Tabletop setup with mounted eye tracker; seen from the side.

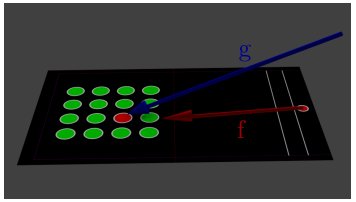

Figure 4: Schematic representation of the information available from flicking and gaze. Line g represents the user's gaze, line f the flicking direction.

Design and Implementation

The most common use-case for a static eye-tracker is to be placed beneath the vertical screen of a desktop computer (or an equivalent setup). Thus finding a configuration that allows the tracking of a large area on a horizontal tabletop surface is a complex task. For one, the geometry of the overall setup is very different. Instead of being positioned centered and in front of the tracked area, looking directly at it, the user is positioned at the side of the surface, which leads to a much lower angle of incident for his gaze. Also the active area of the tabletop is much larger than the common screen size. Because of these differences to the standard setup, a special hardware solution had to be build to allow seamless eye-tracking during touch-based tabletop interaction. The solution proposed in this paper is to mount the eye-tracker head-first above the tabletop surface slightly below the user's eye height (see Figures 1 to 3). From there the user's gaze can be tracked in a large area on the far side of the tabletop, without occluding the active surface area. In addition this setup could be extended to use multiple perpendicular eye trackers to track multiple users on different sides of the table.

A *Session Desk*[1] multi-touch table is used as interactive tabletop. The dimensions of the active surface are $72cm \times 115cm$. The height of the surface is $100cm$. The displayed image had a resolution of 1280×800, providing about $30ppi$. The tabletop is sufficiently large to create a realistic scenario of the intended use case: The user is unable to reach all of the active surface without moving around the table. Also the viewing angles on the surface are as they would be expected to be in a real world scenario.

For eye tracking we employed a Tobii X 60[2], a static binocular eye tracker that produces data with a rate of 60Hz. This eye tracker allows relatively high freedom of movement for the user due to the use of 3D head tracking. There is a $44 \times 22 \times 30cm$ head movement box in which accurate tracking is supposed to be possible. This allows to track users reliable while free-standing and should permit some upper body movement during execution of gestures. The Tobii X 60 calibration is also highly configurable and allows almost arbitrary screen configurations, which allows for native support of the used setup.

The graphical interface as well as the touch interaction was implemented using libavg[3]. Eye-tracking data was accessed through the Tobii SDK 3.0 RC 13[4].

The actual interaction by the user happens through the simple flicking gesture as described in [5]. The gesture mimics the interaction with a physical object that is slid across a plane surface. It is very simple, can be quickly executed and nonetheless provides additional information about the user's intent. The gesture also implies a targeting process which will lead the user's eye to the target. This facilitates the target acquisition by eye tracking without need for explicit eye interaction from the user. The gesture yields yields a line f in 3D space running through the table surface, which corresponds to the xy-plane. In addition from eye and gaze position at the time of selection an additional line g is computed. Ideally both these lines would individually intersect with

[1] http://www.archimedes-exhibitions.de/

[2] http://www.tobii.com/en/eye-tracking-research/global/products/hardware/tobii-x60x120-eye-tracker/

[3] https://www.libavg.de

[4] http://www.tobii.com/eye-tracking-research/global/products/software/tobii-software-development-kit/tobii-software-development-kit/

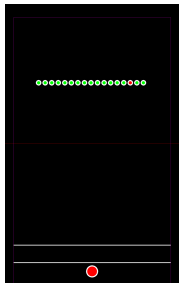

Figure 5: Example of a flicking-task (F) trial.

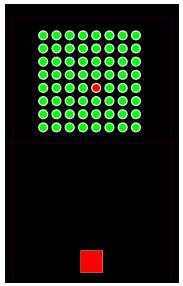

Figure 6: Example of a gaze-task (G) trial.

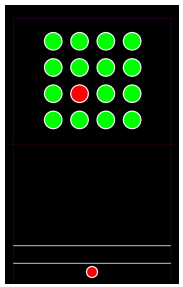

Figure 7: Example of a combi-task (C) trial.

the desired target (see Figure 4). Since this case can not be assumed, and the goal is to use both sources of information to yield more accurate results, the line h, which minimizes the distance between f and g, is computed. Then the closest object (in 3D space) to that line is determined and selected.

Evaluation

A study was performed to evaluate the combination of flicking and gaze for distant target selection in comparison to other selection techniques. The study was conducted using the apparatus described above. A within-subject design was used with three task conditions: Flicking-task (F), gaze-task (G) and combi-task (C). For better comparability of C with the other tasks, it contained a balanced amount of trials using a single row of targets (C_l) or multiple rows (C_f). All tasks were presented in counterbalanced blocks to prevent ordering effects.

Participants

Overall 18 participants (six female), aged 21 to 34 ($M = 26.5$, $SD = 3.67$), successfully participated in the experiment. 6 additional participants had to be excluded from evaluation due to tracking failure. All participants had normal vision and were right handed. The subjects' height ranged between $165cm$ and $186cm$ ($M = 176.63$, $SD = 6.14$).

Procedure

The experiment was conducted in sessions of about 45 minutes. At the start of the session the subjects were asked to fill out a first questionnaire asking for general demoscopic data. Before the experiment started a multi-touch game was presented to familiarize the subjects with multi-touch interaction and our tabletop. The game was a puzzle game that incorporated the flicking gesture and several other multi-touch techniques. Subjects were encouraged to try the different techniques and played for five minutes. Then the subjects were instructed about the experimental procedure and an initial 10-point calibration was performed for the eye tracker.

Each of the trial blocks started with a description of the selection technique, and if necessary a recalibration. A block consisted of 10 practice trials and 40 real trials. After every trial subjects received feedback about their selection; either positive, negative or neutral if the selection was wrong but close. All real tasks were awarded with a score of 10, 0, and -5 for correct, close and wrong answers, respectively. The awarded score and the current overall score were displayed after each trial. The blocks were separated by five minute breaks.

Each task required the subject to select one distinct circle from a group of circles displayed at the far end of the table, as can be seen in Figures 5 to 7. These circles varied in size, spatial arrangement and number between trials. The diameters varied between $10px$, $20px$, $30px$ and $40px$. The items were always presented either in a single row (for F and C_l) or a square field (for G and C_f). This was done to accommodate the accordances of the different tasks, i.e. since the basic flicking gesture is only able to select directional, it is not possible to select a single target from a field of items. Targets were never presented in the outer rows or columns. The number of displayed objects varied according to size to always fill the given area evenly spaced with objects half their diameter apart. The field area was always $500px \times 500px$; a single row had a length of $500px$. Fields and single rows were centered around the same point about $90cm$ from the subject. In each trial the selection results, i.e. correctness and actual selection, were recorded.

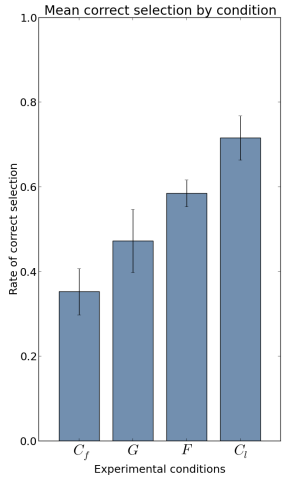

Figure 8: Error rates for each of the conditions C_f, G, F and C_l.

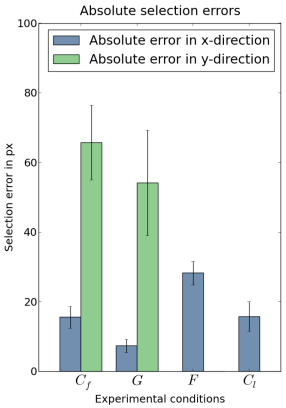

Figure 9: Absolute directional errors for each of the conditions C_f, G, F and C_l.

Results

In our evaluation we were especially interested in the correctness of the selections, therefore we evaluated the general rate of correct selections and the types of error that were made. The reported statistics were acquired using paired t-tests. All graphs show mean values; error bars indicate the standard error of the mean.

Error rates

The error rates achieved for selection in each task can be seen in Figure 8. Significant differences between error rates can be found between C_l and C_f ($F(34) = -5.11$, $p < 0.01$). There is also an improvement of C_l over F ($F(34) = -2.28$, $p < 0.05$) and G ($F(34) = 2.84$, $p < 0.01$). No other significant effects were found.

Directional errors

To further investigate the reasons for these results we looked at the type of selection error that was made. The selection errors are divided in Δ_y, the absolute error in y-direction, and Δ_x, the absolute error in x-direction. The result can be seen in Figure 9. For F and C_l no data for Δ_y is available, since in these condition only single rows were displayed. The differences between Δ_x for all of the conditions, except between C_l and C_f, are significant ($F_{F,C_f}(34) = 2.93$, $p <= 0.01$; $F_{F,G}(34) = 5.72$, $p <= 0.01$; $F_{F,C_l}(34) = 2.44$, $p <= 0.05$; $F_{G,C_f}(34) = -2.42$, $p <= 0.05$). In order from best to worst: G, $C_{l/g}$ and F.

Observations

Subjects were observed to be using two different strategies for flicking: The first and most common way was to start the flicking gesture at the initial position of the throwable. The second strategy was to move the throwable horizontally and start a purely vertical flicking gesture from there. We also observed that a few subjects showed more upper body movement during the flicking gesture than expected. Instead of just moving their arm to execute the gesture, they heavily used their shoulder and upper body during flicking. After the experiment some subjects reported, that during flicking they rarely where looking at the target, but at the throwable instead.

Discussion

We found that C_l had a higher rate of correct selection than flicking (F) or gaze selection (G). This shows that in principle it is possible to achieve the desired synergy effect between multi-touch and gaze input. Unfortunately we did not find this improvement for C_f. The main difference between C_l and the worse performing C_f lies in the geometric arrangement of objects. Inspecting the directional error components we could verify that this is where there is a major difference. The y-component, that is absent in C_l, has an error rate that is much higher than the x-component in both conditions. This high error rate for the y-direction might indicate some particular problem with our current setup. Possible explanations could be a general inaccuracy that occurs due to the large distance or movement of the subject. But this seems unlikely to be the only reason, since this would also affect the x-direction. More probably this effect occurs due to the low viewing angles. Small eye movements and inaccuracies will translate to large changes in the projected position on the table surface, leading to poor accuracy in the y-direction. These problems could be solved by increasing the accuracy of the gaze tracking, or by changing the setup to compensate for the viewing angle.

We also found evidence that there are different strategies employed by our subjects. This leads to difficulties with our simple approach of using the gaze at the time of selection. While it worked reasonable well for many of our

subjects, it cannot be assumed that this approach would work well enough in a scenario "in the wild" where the subjects are complete unaware of the eye tracking. To remedy this problem further data of gaze patterns of subjects, possibly completely unaware of the eye tracker's role for interaction and unaffected by restrictions placed upon them for the sake of accurate eye tracking, should be gathered and analyzed to determine an optimal algorithm to detect the targeted object. This task would preferably be handled in a experiment involving a mobile eye tracker and only a gaze independent flicking task, to gather uninfluenced data.

Conclusions and Future Work

In this work we present the design and evaluation of an interaction technique that relies on gaze and gestural multi-touch input for the selection of distant objects. We propose a multi-touch tabletop equipped with a static eye-tracker as a setup to track users in front of a tabletop and allow gestural and gaze-based interaction. An interaction technique that aims to improve the distant selection of objects by combining gaze and flicking gestures is proposed and evaluated in an user study.

The error in y-direction needs to be investigated as a next step. Furthermore other ways to track gaze data need to be evaluated. We plan a follow-up study with a mobile interactive eye-tracker to gather and analyze flicking and other gestural touch data to determine an optimal algorithm to detect the targeted object.

Acknowledgements

We thank Antti Oulasvirta for valuable feedback. This research project is partially supported by the Nuance Foundation.

References

[1] Abednego, M., Lee, J.-H., Moon, W., and Park, J.-H. I-grabber: expanding physical reach in a large-display tabletop environment through the use of a virtual grabber. In *Proceedings of ITS* (2009).

[2] Bader, T., Vogelgesang, M., and Klaus, E. Multimodal integration of natural gaze behavior for intention recognition during object manipulation. *Proceedings of the ICMI* (2009), 199.

[3] Bragdon, A., and Ko, H.-S. Gesture select: acquiring remote targets on large displays without pointing. In *Proceedings of CHI* (2011).

[4] Jacob, R. J. K. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of CHI* (1990).

[5] Reetz, A., Gutwin, C., Stach, T., Nacenta, M., and Subramanian, S. Superflick: a natural and efficient technique for long-distance object placement on digital tables. In *Proceedings of GI* (2006).

[6] Stellmach, S., and Dachselt, R. Look & touch: gaze-supported target acquisition. In *Proceedings of CHI* (2012).

[7] Stellmach, S., Stober, S., Nürnberger, A., and Dachselt, R. Designing gaze-supported multimodal interactions for the exploration of large image collections. *Proceedings of NGCA* (2011), 1–8.

[8] Turner, J., Bulling, A., and Gellersen, H. Combining gaze with manual interaction to extend physical reach. *Proceedings of PETMEI* (2011), 33.

[9] Ware, C., and Mikaelian, H. H. An evaluation of an eye tracker as a device for computer input. *Proceedings of CHI* (1986), 183–188.

[10] Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (magic) pointing. In *Proceedings of CHI* (1999).