A Field Study on Spontaneous Gaze-based Interaction with a Public Display using Pursuits

Mohamed Khamis
Media Informatics Group
University of Munich
Munich, Germany
mohamed.khamis@ifi.lmu.de

Florian Alt
Media Informatics Group
University of Munich
Munich, Germany
florian.alt@ifi.lmu.de

Andreas Bulling
Perceptual User Interfaces Group
Max Planck Institute for Informatics
Saarbrücken, Germany
bulling@mpi-inf.mpg.de

UbiComp/ISWC '15 Adjunct, September 7–11, 2015, Osaka, Japan.
 ISBN 978-1-4503-3575-1/15/09...$15.00.
DOI: http://dx.doi.org/10.1145/2800835.2804335

Abstract
Smooth pursuit eye movements were recently introduced as a promising technique for calibration-free and thus spontaneous and natural gaze interaction. While pursuits have been evaluated in controlled laboratory studies, the technique has not yet been evaluated with respect to usability in the wild. We report on a field study in which we deployed a game on a public display where participants used pursuits to select fish moving in linear and circular trajectories at different speeds. The study ran for two days in a busy computer lab resulting in a total of 56 interactions. Results from our study show that linear trajectories are statistically faster to select via pursuits than circular trajectories. We also found that pursuits is well perceived by users who find it fast and responsive.

Author Keywords
Pursuits; Smooth Pursuit Eye Movement; Field study; Pervasive displays; Public Displays

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction
Despite ongoing research in eye tracking, gaze-based interaction for everyday use has received little attention as

of today. To close this gap, researchers recently explored ways to incorporate eye tracking into daily interactions.

Gaze-based interaction has the potential to provide numerous benefits to the user and holds particular promise for public displays [8]. Gaze is intuitive [20], fast [18] and natural to use [22]. However, eye tracking researchers face a trade-off between accuracy and usability: in order to collect fine-grained gaze data, each user must go through a calibration process [4] which is, in general, perceived as a tedious task of low usability [10, 27].

While this is acceptable in a desktop setting, given that users are usually engaged for an extended period of time, calibration poses a significant challenge in public space. For example, research in pervasive displays has shown that interaction with screens deployed in public often lasts for just a few seconds, hence requiring “immediate usability” [7, 13] which is challenging to achieve if a calibration process is required.

A possible solution for this is the so-called *pursuits* method [22]. Instead of utilizing fixations or saccades, the *pursuits* technique leverages the smooth pursuit eye movements, which are performed when the eyes follow a moving object [21]. Unlike classical eye tracking techniques, *pursuits* does not determine the absolute gaze point, but instead relies on measuring the correlation between movements of the eyes and movements of dynamic objects on the display. The object whose trajectory correlates most with that of the eye movement, is then determined to be the one the user is looking at [23]. Since it does not rely upon the exact gaze position, the *pursuits* method does not require calibration. Hence it promises fast and spontaneous gaze-based interaction in everyday settings.

The *pursuits* algorithm has been comprehensively analyzed in controlled settings from many aspects. While controlled lab studies have the advantage of isolating external influences, ensuring optimal conditions for the equipments and handling privacy issues (e.g. asking for a participant’s consent to take photos or record videos), they provide low ecological validity and exclude real world dynamics [7, 1]. On the other hand, in-the-wild field studies have the advantage of studying how people unaidedely interact with the system in question. They also allow researchers to investigate aspects such as social effects [12] and audience behavior [14]. Being a method that is meant to offer spontaneous eye-based interaction, it seems plausible to experiment with the method when deployed on a pervasive public display in an in-the-wild field study, to find if it is really welcomed by the passersby. A future step would then be a deployment-based study [1], which is a longitudinal study where the public display deployment is iteratively improved based on user feedback over a long period of time.

We report on the findings of a deployment in a public setting that investigated the effects of the moving object’s speed and trajectory type as well as the time needed to perform a *pursuit* selection. Furthermore, we summarize our observations during the deployment and report qualitative feedback from participants to learn about their experiences when using the novel interaction method.

Our results show that linear trajectories are statistically faster to select via *pursuits* than circular trajectories. We also found that *pursuits* is well perceived by users who find it fast and responsive.

The contributions of this paper are threefold: (1) we describe the setup and execution of an in-the-wild study of pursuits. (2) We report on our analysis of the effects of speed and trajectory of the moving object on the user’s selection speed. (3) We summarize observations and results of semi-structured qualitative interviews with users.

Background and Related Work

We draw from several strands of prior research, most importantly calibration-free eye tracking in general, and smooth pursuits in particular.

Calibration-free eye tracking

Basic gaze-direction estimation has been done using head-tracking and face-detection [3]. More advanced, calibration-free eye tracking methods include relative eye-movement detection, for example the work of Zhang et al. [26, 25]. In their approach, the distance between the center of the pupil and the corner of the eye is calculated, to determine the area at which the user is gazing. Other calibration-free techniques include gaze-gesture detection. For example, work by Vaitukaitis and Bulling [19] detected gaze gestures in different directions using a front-facing camera of a smartphone. Works by Nagamatsu et al. [15, 16] enabled calibration-free eye tracking by using multiple LEDs and cameras. Other researchers proposed simplifying the calibration process. Work by Xiong et al. [24] used an RGBD camera that requires one-time calibration. Pfeuffer et al. [17] relied on the eye's smooth pursuit movement to achieve easier calibration.

All of the aforementioned methods used either video-based or infrared pupil-corneal reflection tracking methods. Another tracking method is the electrooculography-based tracking which measures the electrooculogram (EOG) originating from the eye [5, 10]. Although EOG needs no calibration, currently it requires users to attach electrodes on their skin, making it unsuitable for everyday interactions.

Smooth pursuits

In addition to utilizing smooth pursuits for calibration [17], the same eye movement can be used for explicit interaction. This was first introduced by Vidal et al. [22]. Since its introduction, *pursuits* has been used in several applications ranging from text entry [9], PIN-code entry[6] and entertainment applications [22, 23].

The advantage of pursuits over many other calibration-free techniques is that it also allows high-fidelity interaction. This means that there is a wider range of actions that can be done using pursuits mainly because of the feasibility of showing several *pursuitable* objects.

The effects of the number, speed, and trajectory of moving objects and the correlation parameters on the detection performance have been thoroughly studied in controlled lab settings before [22]. However, an investigation in the wild is still missing as of today. In addition, the effects of speed and trajectory of moving objects on the time required by users to perform a *pursuit* selection have not been subject to research before.

Concept and Implementation

For the purpose of our investigation, we implemented a game that uses *pursuits* as its only input mechanism. In this section we describe the game and the technical parameters we used to implement the pursuits detection algorithm.

The Eye Fishing Game

The game was developed using Java's swing library. The theme of the game is about fishing. The display shows an underwater scene[1] and displays fish moving at different speeds in linear and circular manner. Players are expected to follow the fish with their eyes to catch the fish. Caught fish would then fade away and the game would either proceed to the following level or show the player's score.

[1] CC BY Image by Rafael on Flickr. https://www.flickr.com/photos/rafipics/7914334878

Figure 1: The Eye Fishing game deployed on a 42 inch public display equipped with a Tobii REX. The idle page guides the passersby to step into a green marker on the floor to be in range of the eye tracker and start interacting.

The default idle page (Figure 1) shows the game's title and a message indicating whether or not eyes are detected. Furthermore, a call-to-action label was used to guide the passersby to step into a green marked area on the floor in order to be in range of the eye tracker and start interacting.

Once a user's eyes are detected, the game shows a 4-seconds timer (Figure 2A), during which users are instructed to catch the fish by following it with their eyes.

The game then shows a fish (Figures 2B and 2C). Successfully catching a fish results in the fish fading away and makes another fish appear. After eight fish are successfully selected, the game shows the user's score by displaying the time taken to select all fish (Figure 2D). The game then resets to the idle state, from which a new game begins in case the user is still in range.

Pursuits Detection Algorithm

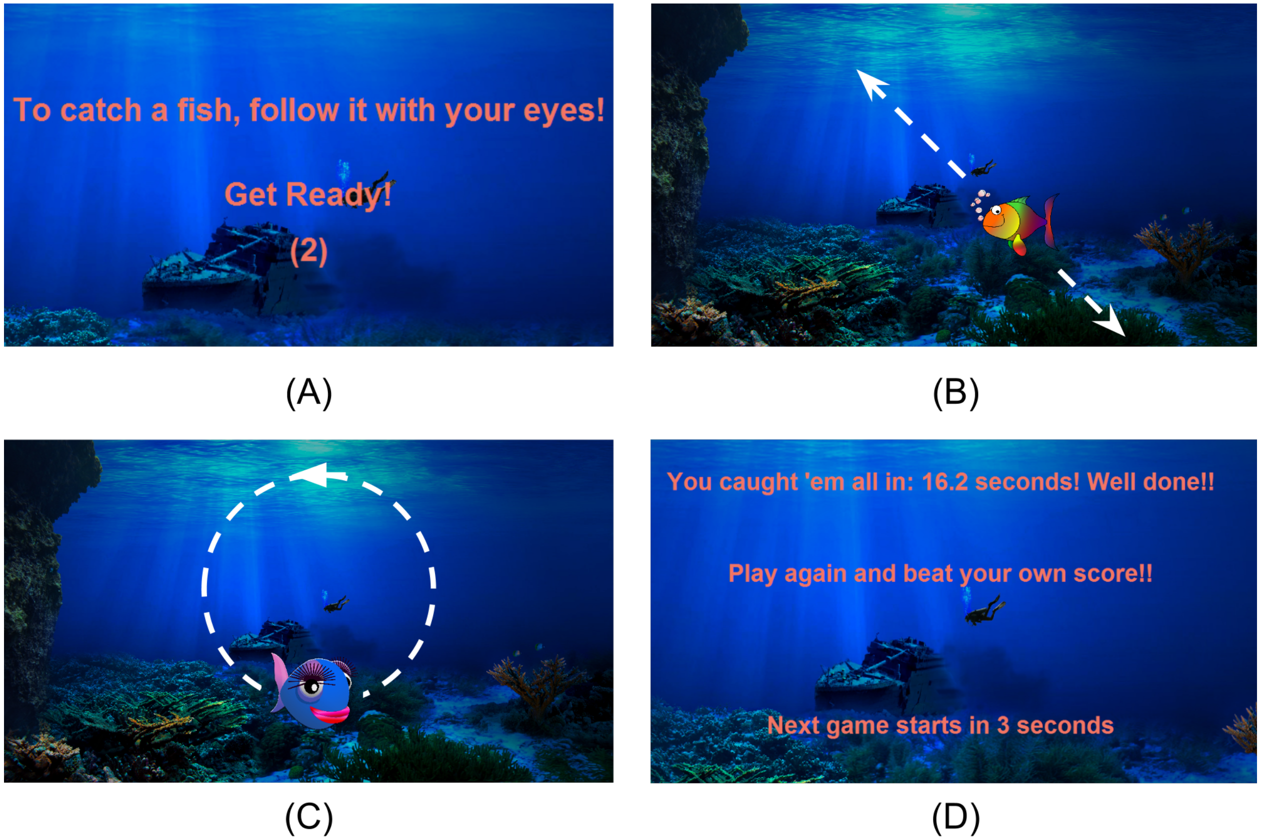
Given a window size (*ws*) and a Threshold (th_{corr}), the algorithm performs a comparison by checking for a correlation between the eye's movements and the fish in a way similar to previous work by Vidal et al. [22]. We used the Pearson's product-moment correlation coefficient[2] to determine the correlation between movements of the user's gaze and the moving fish.

System Parameters

Our system performs a check for correlation using a *ws* of 500ms. That is, every 500ms, the coordinates of the user's gaze and those of the moving fish that were collected within the last 500ms are compared. The resulting correlation is compared against th_{corr} which we set to 80%. This means that the system deduces the user is looking at an object only if the correlation between the movement of that object and that of the eye is higher than 0.8. These values were chosen based on pre-experimentation and previous work by Vidal et al. [22], which showed that high detection rates were achieved using a *ws* of 500ms and high thresholds.

Each game shows eight fish; four follow a circular trajectory, while the other four perform a linear trajectory. For each trajectory type there are two *fast* fish and two *slow* fish. High detection rates of smooth pursuits were reported when objects moved 650 and 450 pixels per second [22]. In our setup, these values correspond to 12.25° and 8.5° visual angle per second respectively. We used these two values for the speeds of *fast* and *slow* fish.

[2]We used the PearsonsCorrelation Class that comes with the Apache Commons Mathematics Library



(A) (B) (C) (D)

Figure 2: A walkthrough the Eye Fishing Game. (A) The loading page instructs the user to follow the fish with his/her eyes. (B) A fish moving in a linear trajectory. (C) A fish moving in a circular trajectory. (D) The recap page shows the user's score and a timer before the next game starts.

Evaluation

Goals

In the context of an in-the-wild deployment of a pursuits-enabled display, the goal of this experiment was to study the effects of (1) the type of the trajectory (linear or circular) and (2) the speed of the moving fish (fast or slow), on the time taken by users to perform the *pursuits* selection. Another goal was to observe participants and collect qualitative feedback to learn about their experiences when using this novel interaction method.

Apparatus

A 42 inch display (3810×2160 pixels) was equipped with a Tobii REX eye tracker (30Hz) and was deployed in an often busy computer lab that is open to university students (Figure 1). Markers were placed on the floor to guide the participants into the eye tracker's range (70cm from the display).

Participants

In total there were 56 interactions with the display, out of which 38 were full-game interactions, in which a participant selected all 8 fish. Twelve participants were interviewed.

Due to the nature of in-the-wild studies it was challenging to collect accurate information about the exact number and demographics of participants. Consecutively played games could have been the result of several participants playing after one another, but it could also be that a participant played multiple times. However being deployed in a university lab we can expect that the majority of the participants were students aging between 18 and 30 years.

Procedure

We deployed the display for two days in a busy computer lab, the game was advertised on social media where it was announced that a new display was installed at which users can catch fish with their eyes. Since it was an in-the-wild deployment, no researchers were present during the entire experiment time, but instead we visited the lab every while to observe and perform semi-structured interviews with participants whom we saw interacting with the system. We asked the participants to describe their experience and indicate the perceived responsiveness of the system (5-point likert scale; 1=Very slow; 5=Very fast).

Design

The study was designed as a repeated measures experiment where all participants were exposed to all conditions. The independent variables of the study were the trajectory type (linear or circular) and the movement speed (*fast* or *slow*), leading to $2 \times 2 = 4$ conditions. At each new game, 8 fish were displayed consecutively one at a time. Every two fish covered one of the conditions. The order of the fish was randomized for each game. Thus, by completing a game, a player would have selected two fish from every condition in a random order.

Measures

During interaction, we logged the times at which games started and ended. We also logged the time at which a fish was selected, along with its trajectory type and its speed. The selection time was calculated starting at the moment the moving fish appeared till the moment it was selected by the user.

Results

When analyzing the results, we excluded all dropouts. This means that we analyzed $8 \times 38 = 304$ pursuit selections.

Pursuit Selection Time

The selection time was calculated starting from the moment the fish appears, till the moment the fish was selected by the user. The window is then cleared and starts again once

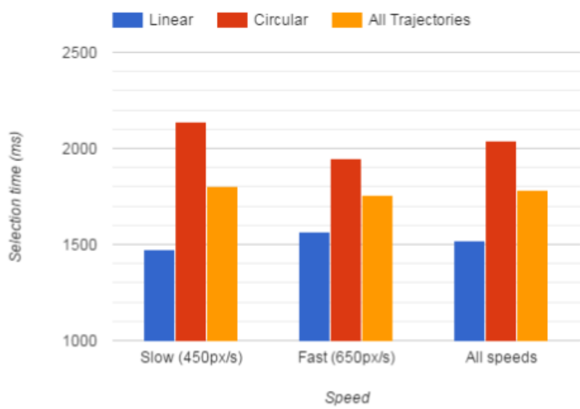


Figure 3: The figure shows that the average pursuit selection time is faster when objects move in a linear trajectory than when objects move in a circular trajectory.

another fish appears. The time taken to perform the fade-out animation was not included in the analysis. A one-way repeated measures ANOVA showed significant main effects for trajectory type on selection time ($F_{1,37} = 10.618$, $p < 0.05$). Post-hoc analyses using Bonferroni correction revealed that there is a significant difference ($p<0.05$) in selection time between linear trajectories (M=1.5, SD=1.3) and circular (M=2.0, SD=1.6) trajectories. This shows that, as illustrated in Figure 3, *pursuit* selection time is significantly faster when using linear trajectories than when using circular ones. However, no significant main effects were found for fish speed on selection time. This can be attributed to the large display size, which made the difference in speeds between 650 px/sec (M=1.76, SD=1.3) compared to 450 px/sec (M=1.8, SD=1.6) insignificant.

Number of Windows

Since the check for correlations happens every 500ms (recall that the selected window size was 500ms), the selection time should optimally be a multiple of *ws*. However, due to processing time, the reported time needed to select is usually few milliseconds more than a multiple of *ws*.

By looking into the number of correlation checks that happened before reaching the threshold th_{corr}, we found that it takes less windows to achieve th_{corr} in the case of linear trajectories (Median=3*ws*) than for circular trajectories (Median=4*ws*). Within the same trajectories, the median of number of checks was the same across different speeds.

Observations

We noticed that participants are more likely to approach the display in groups. The following sequence repeated at least 3 times with different groups: a group passes near a display, one person notices the display and starts interacting, the others then take turns to compete for higher scores. This is similar to the honeypot effect reported in several in-the-wild deployments of public displays [7, 14], where a passerby's interaction encourages others to interact.

It seems that some participants are skeptical to interact if they are alone. In two cases, participants noticed the display, but only interacted after calling others to join.

Participants get frustrated quickly when the system takes more time to respond. An observed participant was very dissatisfied when the fish was not selected despite following it. This aligns with previous work in interaction with public displays [11], which showed that even the slightest delay when interacting with public displays is problematic and can lead to frustration and abandoning the display.

It was noticed that taller participants had to lean to be recognized by the system. A participant who was accompanied by his son had to carry his son in order to be recognized by the eye tracker.

Interviews

Out of the 38 full-interactions, we interviewed 12 participants (3 females). Overall the interviewed participants reported that the interaction is well perceived. They find interaction via gaze to be interesting, fast and easy. The system was indicated to be responsive (M=3.5, SD=0.8).

When asked, ten out of twelve interviewed participants indicated that they noticed the different trajectories. However none reported any difference in perceiving the different trajectories. One participant noticed that some fish are faster than others. She found the faster objects easier to follow, while slower ones felt boring and unnatural.

Discussion

Although gaze is a relatively new technology that our participants are less likely to have used before, it is impressive that, despite minimal instructions, the system could be used easily by our participants. We attribute this to the nature and intuitiveness of gaze and smooth pursuits, and the feasibility of using it without prior training.

The results show that interaction via gaze using *pursuits* is responsive and well-perceived. This shows that *pursuits* is suitable for public display deployments.

The results also show that for our setup, linear trajectories are significantly faster to select by users (1.5 seconds) than circular trajectories (2.0 seconds).

The observations drew attention to the fact that public displays require immediate usability. This shows that public displays offer a challenging yet realistic testing ground for usable state-of-the-art gaze-interaction mechanisms.

We attribute the inability of some participants to notice the different trajectories to the fast selection speed, which did not allow the participants to notice a difference in trajectory. Consequently, it is advised to use a higher threshold and window size values in cases where such a delay is desirable (e.g. to give a chance for examining the dynamic object before selecting it). It should also be noticed that despite the fast nature of the human gaze, any delay in the system response is not welcomed by the passersby who are usually not intending to spend a lot of time at the public display.

Among the observations, we noticed that participants were skeptical towards interacting with the display when they are alone, but more willing to interact when surrounded by acquaintances. It is not clear whether the observed hesitance to interact alone is due to the interaction method or due to the nature of the application. Further work can investigate different applications to identify whether or not the interaction modality is embarrassing to use in public.

Limitations and Future Work

Future experimentation may try studying the impact of the number of moving objects on the selection speed. The absence of user-dependent calibrated gaze-data makes it challenging to identify the moment the user starts to follow an object in real-time. A work-around is to first calibrate the eye tracker per user. Although this would defeat the main motivation behind using *pursuits*, the fine-grained gaze data can then be used to estimate the user's selection speed in the case of displaying multiple objects.

While testing the game, we noticed that a correlation threshold of 0.95 is reasonable in a desktop setting. However using the same value on the public display makes selection very difficult. Hence, future work should investigate the influence of display size on the threshold.

A challenge that was observed during the study is that eye trackers will have to deal with passersby with different heights. This needs to be addressed for all future in-the-wild deployments of eye trackers. A possible direction is to investigate ways of automatically detecting the user's height and adjusting the eye tracker's angle accordingly. We note however, that an eye tracker with a larger range could detect taller players if they stand farther from the display.

We used object speeds similar to those reported by Vidal el al. [22]. However because speeds could be perceived differently across displays with different resolutions, we recommend future deployments to define speed as a factor of the display's size or in visual angle per second.

Future work can investigate trajectory types in further detail. For example, studying the effects of manipulating the angle at which the linear trajectory moves and the effects of zigzag or irregularly shaped movements on user experience and selection speed.

It was shown that many users do not like to interact using mid-air gestures in public space due to social embarrassment [2]. We were surprised that some participants were hesitant towards interacting alone with our application, given that gaze is a subtle interaction method. This should be further investigated to verify whether or not gaze causes social awkwardness.

Conclusion
In this paper we reported on an in-the-wild field experiment where we deployed a *pursuits*-enabled display in a public space. Our results indicate that *pursuits* selection time is significantly faster in the case of linear trajectories compared to circular trajectories. We showed that participants perceive the interaction method positively and find it fast and responsive.

References

[1] Florian Alt, Stefan Schneegaß, Albrecht Schmidt, Jörg Müller, and Nemanja Memarovic. How to Evaluate Public Displays. In *Proc. PerDis '12*. ACM, Article 17, 6 pages.

[2] Harry Brignull and Yvonne Rogers. 2003. Enticing people to interact with large public displays in public spaces. In *Proc. INTERACT '03*. 17–24.

[3] Frederik Brudy, David Ledo, Saul Greenberg, and Andreas Butz. Is Anyone Looking? Mitigating Shoulder Surfing on Public Displays Through Awareness and Protection. In *Proc. PerDis '14*. ACM, 1–6.

[4] Andreas Bulling, Florian Alt, and Albrecht Schmidt. Increasing the Security of Gaze-based Cued-recall Graphical Passwords Using Saliency Masks. In *Proc. CHI '12*. ACM, 3011–3020.

[5] Andreas Bulling, Daniel Roggen, and Gerhard Tröster. 2009. Wearable EOG Goggles: Eye-based Interaction in Everyday Environments. In *CHI '09 EA*. ACM, 3259–3264.

[6] Dietlind Helene Cymek, Antje Christine Venjakob, Stefan Ruff, Otto Hans-Martin Lutz, Simon Hofmann, and Matthias Roetting. 2014. Entering PIN codes by smooth pursuit eye movements. *Journal of Eye Movement Research* 7(4):1 (2014), 1–11.

[7] Nigel Davies, Sarah Clinch, and Florian Alt. 2014. *Pervasive Displays - Understanding the Future of Digital Signage*. Morgan and Claypool Publishers.

[8] Mohamed Khamis, Andreas Bulling, and Florian Alt. Tackling Challenges of Interactive Public Displays using Gaze. In *Proc. UbiComp '15 Adjunct*. ACM.

[9] Otto Hans-Martin Lutz, Antje Christine Venjakob, and

Stefan Ruff. 2015. SMOOVS: Towards calibration-free text entry by gaze using smooth pursuit movements. *Journal of Eye Movement Research* 8(1):2 (2015), 1–11.

[10] Päivi Majaranta and Andreas Bulling. 2014. *Eye Tracking and Eye-Based Human-Computer Interaction*. Springer Publishing, 39–65. DOI: http://dx.doi.org/10.1007/978-1-4471-6392-3_3

[11] Paul Marshall, Richard Morris, Yvonne Rogers, Stefan Kreitmayer, and Matt Davies. Rethinking 'Multi-user': An In-the-wild Study of How Groups Approach a Walk-up-and-use Tabletop Interface. In *Proc. CHI '11*. ACM, 3033–3042.

[12] Nemanja Memarovic, Ivan Elhart, and Marc Langheinrich. FunSquare: First Experiences with Autopoiesic Content. In *Proc. MUM '11*. ACM.

[13] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. Requirements and Design Space for Interactive Public Displays. In *Proc. MM '10*. ACM, 1285–1294.

[14] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. Looking Glass: A Field Study on Noticing Interactivity of a Shop Window. In *Proc. CHI '12*. ACM, 297–306.

[15] Takashi Nagamatsu, Kaoruko Fukuda, and Michiya Yamamoto. Development of Corneal Reflection-based Gaze Tracking System for Public Use. In *Proc. PerDis '14*. ACM, Article 194, 2 pages.

[16] Takashi Nagamatsu, Ryuichi Sugano, Yukina Iwamoto, Junzo Kamahara, and Naoki Tanaka. User-calibration-free Gaze Tracking with Estimation of the Horizontal Angles Between the Visual and the Optical Axes of Both Eyes. In *Proc. ETRA '10*. ACM, 251–254.

[17] Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible. In *Proc. UIST'13*. ACM, 261–270.

[18] Linda E. Sibert and Robert J. K. Jacob. Evaluation of Eye Gaze Interaction. In *Proc. CHI '00*. ACM, 281–288.

[19] Vytautas Vaitukaitis and Andreas Bulling. Eye Gesture Recognition on Portable Devices. In *Proc. PETMEI '12*. 711–714.

[20] Roel Vertegaal. 2003. Attentive user interfaces. *Commun. ACM* 46, 3 (2003), 30–33.

[21] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. Detection of smooth pursuits using eye movement shape features. In *Proc. ETRA '12*. 177–180.

[22] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proc. UbiComp '13*. ACM, 439–448.

[23] Mélodie Vidal, Ken Pfeuffer, Andreas Bulling, and Hans Gellersen. Pursuits: eye-based interaction with moving targets. In *Proc. CHI '13*. 3147–3150.

[24] Xuehan Xiong, Qin Cai, Zicheng Liu, and Zhengyou Zhang. 2014. Eye Gaze Tracking Using an RGBD Camera: A Comparison with a RGB Solution. In *Proc. PETMEI '14*. ACM, 1113–1121.

[25] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays. In *Proc. CHI '13*. ACM, 851–860.

[26] Yanxia Zhang, Jörg Müller, Ming Ki Chong, Andreas Bulling, and Hans Gellersen. GazeHorizon: Enabling Passers-by to Interact with Public Displays by Gaze. In *Proc. UbiComp '14*. ACM.

[27] Zhiwei Zhu and Qiang Ji. 2005. Eye Gaze Tracking Under Natural Head Movements. In *Proc. CVPR '05*. IEEE Computer Society, 918–923.