Pers Ubiquit Comput
DOI 10.1007/s00779-015-0866-8

ORIGINAL ARTICLE

Eye tracking for public displays in the wild

Yanxia Zhang[1] · Ming Ki Chong[1] · Jörg Müller[2] · Andreas Bulling[3] · Hans Gellersen[1]

Received: 29 October 2014 / Accepted: 12 January 2015

Abstract In public display contexts, interactions are spontaneous and have to work without preparation. We propose gaze as a modality for such contexts, as gaze is always at the ready, and a natural indicator of the user's interest. We present *GazeHorizon*, a system that demonstrates spontaneous gaze interaction, enabling users to walk up to a display and navigate content using their eyes only. GazeHorizon is extemporaneous and optimised for instantaneous usability by any user without prior configuration, calibration or training. The system provides interactive assistance to bootstrap gaze interaction with unaware users, employs a single off-the-shelf web camera and computer vision for person-independent tracking of the horizontal gaze direction and maps this input to rate-controlled navigation of horizontally arranged content. We have evaluated GazeHorizon through a series of field studies, culminating in a 4-day deployment in a public environment during which over a hundred passers-by interacted with it, unprompted and unassisted. We realised that since eye movements are subtle, users cannot learn gaze interaction from only observing others and as a result guidance is required.

Keywords Eye tracking · Gaze interaction · Public displays · Scrolling · Calibration-free · In-the-wild study · Deployment

1 Introduction

Public displays present a challenging interaction context. Passers-by engage with public displays spontaneously, driven by opportunity or emerging information needs. Users and displays come together as strangers, and yet interactions have to work without preparation, as they are usually unplanned. In spite of the challenges, we believe eye gaze to be a modality of significant potential for the public display context. Our motivation is that humans naturally express interest by their gaze. We do this spontaneously when a display "catches our eye". Users need not be in touching distance to express gaze input and can use gaze through window panes. Public display interactions are typically of short duration, which favours gaze as a modality that is always at the ready while also limiting potential fatigue.

Public display applications vary in their need for input. There are many scenarios in which information displays might be improved by gaze control of low complexity. Arrival/departure screens in airports, mounted overhead and out of reach, display content that is split over a number of pages which could be navigated by gaze. Situated map displays could be made gaze responsive to provide more detail on areas of interest. Shop displays could employ gaze to let window-shoppers scroll through offers of

✉ Yanxia Zhang
yazhang@lancaster.ac.uk

Ming Ki Chong
mingki@acm.org

Jörg Müller
joerg.mueller@acm.org

Andreas Bulling
bulling@mpi-inf.mpg.de

Hans Gellersen
hwg@comp.lancs.ac.uk

[1] Lancaster University, Lancaster, UK

[2] Aarhus University, Aarhus, Denmark

[3] Max Planck Institute for Informatics, Saarbrücken, Germany

Published online: 03 July 2015

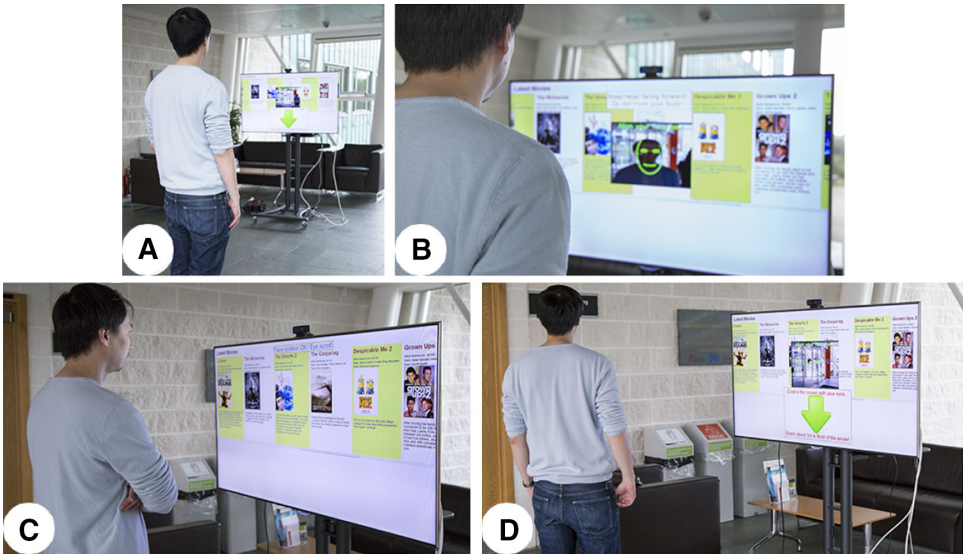


Fig. 1 GazeHorizon enables users to interact with a public display using only their eyes. **a** *Encounter*: a passer-by encounters a GazeHorizon display. **b** *Walk Up and Position*: the user walks up to the display, and the interface shows visual cues to assist the user to position himself. **c** *Eye-based Interaction*: the user navigates horizontal content by moving his eyes; whatever content the user looks at drifts to the centre where it comes to a halt. **d** *Walk Away*: the system resets back to the initial state after the user walks away and is ready for the next user.

interest. These scenarios have in common that they describe spontaneous interactions, where users walk up to a display they may never have seen before and should able to interact without further ado.

Enabling gaze for public displays presents a technical challenge for robust and person-independent tracking of gaze in the wild, and a design challenge for realising gaze control in ways that users can discover and use "on the spot". Existing eye trackers are designed to accurately track user attention on screens but require controlled conditions [4]. In HCI, they have been embraced for gaze pointing, where gaze is used as alternative or complement to other pointing devices [5, 27]. However, gaze pointing requires careful adaptation to individual users who need to complete a calibration procedure prior to interaction, hampering spontaneous use [11]. We argue that public display settings require a different approach to eye tracking and gaze input, less fixated on high fidelity but optimised for instantaneous usability by any user without prior configuration, calibration or training steps.

We present *GazeHorizon*, a system developed to demonstrate spontaneous gaze interaction with public displays [30], illustrated in Fig. 1. The system tackles three significant challenges to facilitate gaze as a walk-up-and-use modality. The first challenge is eye tracking in the wild, for which our system relies on a single camera mounted with the display. We use computer vision to track the face of an approaching user, extract and process eye images, and compute the horizontal gaze direction using the PCR technique [29]. The resulting input is limited to the horizontal dimension as vertical eye movement is not as robustly attainable with a single camera, but it is person independent and obtained without any set-up phase.

The second challenge tackled in GazeHorizon is to map gaze input to intuitively usable control of information displays, in the absence of a rigid calibration. Conventional eye gaze mappings are absolute, where a gaze direction is directly associated with a point on the screen. In GazeHorizon, we use a relative mapping instead, for rate-controlled navigation of horizontally organised information on the screen. The control users gain over the display is limited to navigation along one dimension, but the mapping provides a robust user experience where the gaze control task is naturally aligned with the visual perception of the displayed content.

The third challenge addressed in GazeHorizon is to guide users to discover how they can interact. In conventional use of gaze interfaces, users are either trained a priori or assisted by an expert (e.g. by an experimenter in a usability laboratory). In the public display context, the system itself has to provide all necessary assistance to bootstrap interaction, as users have no prior awareness of the system's interactive capability. To address this problem, we have conducted field studies with the system to understand what guidance users require and to develop visual cues and interactive assistance. The result is a system that enables unaware users to control a display by gaze, without any training or hand-holding.

GazeHorizon was developed and evaluated through iterative field studies.[1] A first study was designed to elicit insight into the guidance needed for users to be able to use gaze for display interaction. A second study tested the

[1] This article is an extended version of the work [30] we presented in UbiComp 2014. Previous work only highlights the deployment of GazeHorizon, whereas this article provides a complete report on detailed observations, lessons learned and in-depth discussions for each field study.

efficacy of visual cues added to the system, by inviting passers-by to complete a task relying solely on the hints provided by the system itself. In a final study, the system was deployed in the wild for 4 days to evaluate how passers-by would interact with it. Over 100 users were observed engaging with the system, unprompted and unassisted. Post-interviews with a random sample of users confirmed that most users had been able to use the system to explore displayed information by gaze.

The contributions of our work are as follows:

1. In the sum of our work, we provide a real-world demonstration of spontaneous gaze interaction with public displays. This represents a first work on gaze with public displays that goes all the way to show that gaze can be leveraged in challenging contexts, where users engage spontaneously and without prior awareness of the gaze capability.
2. The GazeHorizon system itself is novel in the interactive capabilities it demonstrates, and the concepts it employs to achieve these. Innovative aspects include how gaze input is acquired with a single camera (expanded in [29]), how gaze is mapped for intuitive display control and how passers-by are interactively guided to bootstrap interaction.
3. The system has been developed and evaluated through extensive field studies, through which we contribute original insight into challenges in enabling passers-by to use gaze, and the response of users to different strategies embodied in the design of our system and tested in the wild.

2 Related work

Gaze is long established as interaction modality [5], but only little explored for public display settings. The prevailing paradigm for gaze input is pointing with a direct mapping to display content, for instance, for desktop control [27], eye typing [8], enhanced scrolling [7] and remote target selection [21, 22]. This requires accurate tracking of a user's gaze and careful calibration before interaction can commence. In contrast, we demonstrate use of gaze at lower fidelity but optimised for instantaneous use in a public display context. Our approach is person independent which means that there is no calibration to individual users, and it requires only minimal instrumentation with a single camera.

Early work on gaze for settings beyond the laboratory introduced infrared-based eye contact sensors to make displays attention aware [20, 23]. Gaze locking recently explored eye contact detection at-a-distance with a single camera [19]. Beyond eye contact, Vidal et al. [24] showed that content displayed in motion can be gaze selected without calibration by exploiting smooth pursuit eye movement, however, relying on higher fidelity tracking of gaze with specialist hardware. Zhang et al. [28] introduced Sideways, a system that classifies gaze into three directions (left, centre and right). Scrolling was proposed as one application for Sideways, where a user triggers discrete scrolling events by casting glances to the left or right. GazeHorizon is similar in utilising horizontal gaze but employs a relative mapping for rate-controlled navigation of displayed information, which is designed to be aligned with natural gaze attention to content of interest. Other research on gaze with public displays has considered specific techniques for gaze estimation (e.g. [14, 18]) and gaze application (e.g. [3, 21, 22]), but these works abstract from the real-world deployability on which our work is focused.

Previous research found that a major challenge in public display research is to make passers-by aware of the interactive affordances of the display and to entice them to approach the device and begin interaction [6]. However, the design of a gaze interface often starts with the assumption that users are aware of the gaze interactivity. GazeHorizon addresses this problem by integrating interactive guidance and is a first gaze interface for public displays that passers-by are able to use without any prior awareness of how the system works.

While gaze has not been considered much, there is a plethora of work employing other modalities for public display interaction [12]. The most common ones are touch (e.g. CityWall [15]) and attached input devices such as a keyboard (e.g. Opinionizer [2]). For settings where a display is unreachable or cannot be manipulated directly, research has suggested the use of mobile devices for remote interaction (e.g. TouchProjector [1]), and whole-body gesture interfaces to eliminate the need for additional devices [16]. Our focus on gaze, in comparison, is motivated by the flexibility of gaze as a hands-free modality that is always at the ready and usable with displays that are out of reach.

Users behave differently in public in comparison with a controlled setting [10]. For this reason, research in public displays is often conducted in the field to understand people's actual behaviours. Researchers have explored ways to entice people to interact with displays in public [2] and observed how people used a multi-touch display in various social configurations [15]. More recently, researchers explored methods for passers-by to notice interactivity of public displays [13], revelation of specific body gestures for interaction [26], as well as content legibility on very large displays while walking [17]. Yet, no previous research has explored the use of gaze for user interaction with displays in the public domain. GazeHorizon, in contrast, has been developed through a series of field studies to understand the system's use in practice.

3 System design of GazeHorizon

The goal of GazeHorizon is to create a system that allows any passes-by to walk up to a display and to navigate the displayed information using only their eyes. The system is designed for deployability in the wild and requires only a single off-the-shelf web camera—placed either in front or on top of a display—to capture the presence of users (at a frame rate of 30 Hz and a resolution of 1280 × 720 px) and supports interaction of one user at a time. From a system's perspective, the interaction involves the following processes:

Face detection and tracking: For every thirty frames, the system scans for the presence of users looking towards the display, by detecting frontal faces. If a group of users approach the system, only the person standing in the centre of the screen (i.e. the face positioned in the central region) is tracked continuously.

Eye tracking: When the image resolution of the tracked face is larger than 200 × 200 px, the system extracts eye feature points from the face image. The size restraint corresponds to a user standing at a distance of approximately 1.3 m away from the camera.

Gaze interaction: The system uses the extracted eye feature points for computing horizontal gaze direction. The gaze direction is mapped to rate-controlled scrolling of the display. When the user looks to the left, the display is scrolled to the right (and vice versa), and the scroll speed varies based on how far away from the centre that the user looks.

Reset: If the system detects that the user's face has disappeared from its field of view, the system resets back to the initial *Face detection and tracking* phase.

3.1 Estimating horizontal gaze directions

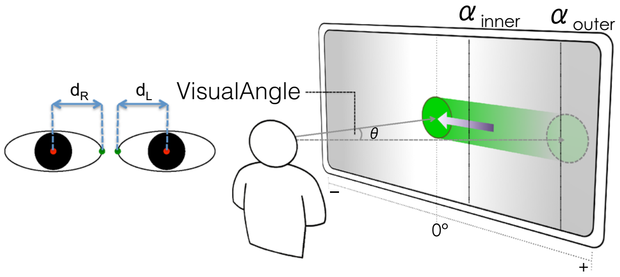


Fig. 2 GazeHorizon uses the distances between the pupil centres from the inner eye corners d_L and d_R for estimating horizontal gaze direction, using the PCR technique [29]. The gaze direction is mapped for rate-controlled scrolling with a self-centring effect. If the user looks at an object on the right, the display will scroll to the left; as the user's gaze naturally follows the object-of-interest, the scrolling speed will decrease and bring the object to a halt in the centre of the display

We use the image processing techniques outlined by Zhang et al. [28] in the *Face detection and tracking* and *Eye tracking* phases to extract pupil centres and inner corners. The system calculates the horizontal distance between the two points for both left (d_{L}) and right (d_{R}) eyes. As illustrated in Fig. 2, d_{L} and d_{R} are relatively equal when a user looks straight ahead. However, if the user looks towards one side, one of the distances increases, while the other decreases. We exploit the ratio of the distances for estimating gaze direction, using the Pupil-Canthi-Ratio (PCR) defined as the distance ratio between d_{L} and d_{R} [29] where:

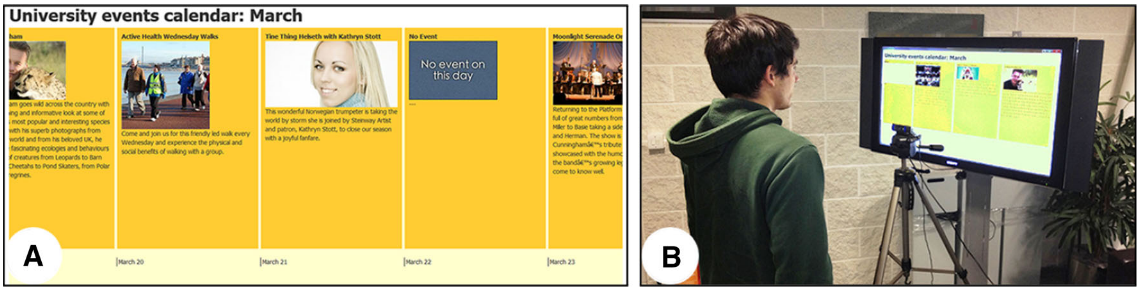
$$PCR = \begin{cases} -(\frac{d_{\mathrm{L}}}{d_{\mathrm{R}}} - 1), & \text{if } d_{\mathrm{L}} - d_{\mathrm{R}} > 0; \\ \frac{d_{\mathrm{R}}}{d_{\mathrm{L}}} - 1, & \text{if } d - d_{\mathrm{R}} < 0; \\ 0 & \end{cases} \tag{1}$$

PCR represents the visual angle of gaze, relative from "looking straight ahead". A negative PCR denotes a look to the left, while positive denotes looking right, and a zero PCR indicates that the user is looking straight ahead.

3.2 Relative mapping of gaze to display control

We use a relative mapping of the estimated gaze direction, as illustrated in Fig. 2. The mapping is designed to leverage implicit navigation of the displayed content, where the scrolling action is triggered by attention to the content and not perceived as an explicit command. We follow the principle of whatever a user looks at it moves to the centre [31]. When the user looks at content located away from the centre of the display, this results in the scrolling effect of moving content towards the centre. If the user follows the content as it becomes centred, the scrolling speed decreases and the content comes to a halt in the centre of the display.

In order to provide a stable display experience, we do not map PCR uniformly to scrolling speed, but define lower and upper thresholds for visual angle relative to the display


University events calendar: March
ham
Active Health Wednesday Walks
Tine Thing Helseth with Kathryn Stott
No Event
Moonlight Serenade Or
No event on this day
March 20
March 21
March 22
March 23
A
B

am goes wild across the country with ing and informative look at some of most popular and interesting species with his superb photographs from world and from his beloved UK, he fascinating ecologies and behaviours f creatures from Leopards to Barn Cheetahs to Pond Skaters, from Polar egrines.

Come and join us for this friendly led walk every Wednesday and experience the physical and social benefits of walking with a group.

This wonderful Norwegian trumpeter is taking the world by storm she is joined by Steinway Artist and patron, Kathryn Stott, to close our season with a joyful fanfare.

Returning to the Platform full of great numbers from Miller to Basie taking a side and Herman. The show is Cunninghamâ€™s tribute showcased with the humc the bandâ€™s growing le come to know well.

Fig. 3 (Field study 1) **a** GazeHorizon event browser interface. **b** An illustration of the study set-up

centre. This allows us to set scrolling speed to zero for a wider region in the centre of the display and to a maximum constant for larger visual angles. Note that the system is based on gaze direction and does not actually predict what the user is looking at. However, the interaction design provides a robust illusion of response to the user's point of regard. We suggest this mapping of gaze is useful for exploration of larger information spaces that can be presented in one dimension along the horizontal axis. A good example is a timeline, of which the display reveals a part at any time. The further the user looks towards the edge of the display, the faster it will scroll to the effect of revealing new information "in the direction of interest".

4 Guidance for public gaze-based interaction

GazeHorizon is designed for spontaneous gaze interaction in the public domain. Having an expert onsite to instruct users is impractical. Instead, displayed information should be self-explanatory and no prior training of usage should be required [25]. Gaze is currently an uncommon modality; many people are unfamiliar with using gaze for interaction. Our aim is to create a stand-alone interface with interactive guidance that (1) communicates interactivity to novice users, so they understand the interaction model of GazeHorizon; (2) provides the users with appropriate guidance, so they autonomously adjust themselves into an optimal position for the system to track their eyes; and (3) gives dynamic real-time feedback according to the actions of the users. We conduct a series of field studies to understand how novice users engage GazeHorizon spontaneously.

5 Field study 1: levels of required guidance

The aim of this study is to understand what information novice users need in order to use GazeHorizon; hence, the level of guidance they require to figure out the interaction model. We thus conducted the study in the field by deploying our system in a public area.

5.1 Setup

We used a 40-inch LCD display (1280 ×768 px), positioned at the height of 115 cm above ground (from the lower bezel). We placed a web camera on a tripod at the height of 150 cm above ground and 50 cm in front of the display centre. We implemented GazeHorizon as an event timeline browser. Passers-by could browse through events happening around the local area (see Fig. 3). Events were sorted in ascending chronological order. Looking left scrolls for earlier events, while looking right scrolls for future events. Each event was displayed in a 300 × 480 px box, with a title, a picture, text description and the date shown at the bottom.

We conducted the study over a 2-day period, in the reception area of a university building. The building hosts a computer science department, a café and numerous technology companies. People with various technology background passed by everyday. The area was illuminated by ambient daylight, and ceiling light was turned on during late afternoon.

5.2 Procedure

During the study, a researcher invited passers-by to take part. The researcher introduced the system as an interactive display that showed local events, but never revealed the system was eye based during the introduction. Thereafter, the participant was encouraged to experience the system.

To test for intuitiveness, we evaluated the amount of instructions (or guidance) that users needed to comprehend the operation correctly. We predefined the instructions into five stages. Table 1 lists the instruction protocol and the number of participants who needed the levels. The researcher started off by inviting participants to stand in front of the display (hence, giving *L1* instruction) and then prompted the participant for their perceived interaction model. If the participant answered incorrectly or required further guidance, the researcher gradually revealed the next level and prompted the participants again. This continued until the participant realised the correct operation or the whole set of instructions was revealed.

Table 1 (Field study 1) Five levels of guidance we provided to our participants, in ascending order

Instruction levels and hints		Count
L1	Stand in front of the display.	10
L2	The system reacts to eyes.	11
L3	Keep head still, face towards the display, and move eyes only.	5
L4	Look specifically at each event object.	0
L5	Look at an event object and follow it. The object stops in the centre.	0
*	*Failed to use the system after all five levels were revealed*	*4*

The count column indicates the number of participants who needed up to that level of instruction to determine the interaction of GazeHorizon

After experiencing our system, we interviewed the participants for qualitative feedback. We first asked a set of predefined questions and then further prompted them with probing questions for detailed explanation.

5.3 Results

In total, we invited 30 passers-by (8 females). Seven participants wore glasses, and five of them removed their glasses during the study. The participant stood at an average distance of 120 cm (SD = 10) away from the display. At this distance, our system captured the participant's face at an average resolution of 245 × 245 px (SD = 40).

5.3.1 Levels of guidance required

In total, twenty-six participants successfully comprehended the operation, and twenty-four of them acknowledged that it was easy to figure out.

Ten participants required only level 1 instruction. They commented that the system was "*easy*", "*self-explanatory*", "*just look at the events... and focus on what I want to read*". They explained that they realised the system was eye-based when they noticed the movement of the content corresponded to the movement of their eyes. For instance, a participant reported that "*[the picture] moved when my eyes moved*".

Eleven participants needed up to the level 2 instruction. They explained that eye-based input was uncommon. For example, a participant said "*I felt something is moving; however, I am not sure what [the system] reacts to.*" Six participants first attempted to touch the screen, wave their hands or use body gestures. They imagined the system was "*motion-based*" because they were aware of existing motion capture technology (e.g. Kinect). However, after we revealed that the system was eye based, the interaction became "*obvious*". The participants mentioned that the level 2 instruction was important, as it eliminated them from attempting to use their body for interactions.

Five participants needed up to level 3 instruction. When they looked at the displayed information, they turned their head to face towards it. After they were told to keep their head still and face towards the centre of the display, the participants realised the system reacted to eye movements.

Four participants failed to use our system. Two of them could not understand the interaction model, even full instructions were given. Post-study analysis revealed that one participant failed because the system failed to detect his pupil centre. Also, one participant declined to retry after a failed attempt.

5.3.2 Users' perceived control

From the interview, seventeen participants perceived that they were in control of the system and it responded accurately. However, eight participants felt that the display information was jittery and the system had delays, which prohibited them to apprehend the event information. Some participants criticised that it was "*unnatural to keep the head still*", because they intuitively turned their head to face the direction to where they paid attention.

We prompted the participants to explain their perceived interaction model of the system. Five people anticipated that more information was beyond the display border, so they fixated on the corner/edge to scroll for more information. The majority realised that a fixation at the centre of the display would stop the scrolling. Five participants acknowledged that the scrolling speed increased as they looked further away from the centre of the display. Eight people responded that they scrolled the timeline by moving their eyes left or right. They comprehended eye movement as a triggering action for scrolling, but neglected the speed difference.

Some participants mentioned that they imagined their gaze as a pointer. When they looked at a displayed object, they expected visual feedback, like "*highlight*" or "*magnification*". Other participants explained that they used the displayed objects (e.g. pictures, titles of events) or the screen edges and corners as visual stimuli for gaze fixation.

5.3.3 Three patterns of user operation

We observed three different ways of how the participants used our application. The majority of the participants' perceived concept was in line with the interaction model we designed. They preferred to read detailed information from the centre of the display. One participant related this to his desktop; he arranged items (e.g. windows and icons) in the centre for a better viewing.

The second pattern was acknowledged by six participants. They sometimes read short sentences immediately as they entered the display from the side. However, a few participants found that "*it was disturbing*" trying to read moving information, which easily caused them to lose focus. Even though the information was moving, as the participants followed it towards the centre, the information slowed down and eventually became stationary. They can then read the information. Also, if users turn their head to look towards the side, the scrolling will halt, because the system could no longer detect the presence of a frontal face looking towards the screen. This inherently allows the participants to read information on the side by turning their head.

The third pattern was identified by two participants. They found that the system was difficult to control, even though they were fully informed of the instructions. They got distracted easily by moving objects, so they often moved their eyes to look for new information on the display. As a result, the interaction model failed because the participants did not fixate on the centre to stop the scrolling. This is essentially the midas touch problem. The scrolling action is triggered unintentionally and causes user frustration.

5.3.4 Alternative interactions discovered by users

Surprisingly, five participants discovered that they can fixate on the centre of the display and slightly turn their head; "*the movement of the picture is synchronised with head movement*". This is similar to Mardanbegi et al.'s head gestures technique [9]. After a few trials, the participants acknowledged that "*[the system] was less responsive to head turning* and they could not focus on what they wanted to see.

Also, a participant suggested that she could stop scrolling by looking downward, such as staring at the event's date at the bottom. Looking downward caused her eyes to be occluded by her eye lids and eye lashes, so the system stopped detecting her eyes.

5.4 Summary

From this study, we found that many people are unaware of gaze as an input modality for large display navigation. Our results revealed that *L2* and *L3* instructions are vital for communicating the gaze interactivity. Thus, we translated these levels into visual labels and embedded the labels in the interface for our next study.

6 Field study 2: testing interactive guidance

Our aim is to design a stand-alone application, where users interpret the interaction solely based on information given on the interface. From field study 1, we learned that users needed three levels of guidance: (1) *position* (stand in front of the display), (2) *eye input* (the system reacts to users' eyes) and (3) *head orientation* (keep head facing forward and move eyes only). We embedded these three levels as visual cues on the interface for instructing users.

Previous research showed that textual information was very effective in enticing interaction on public displays [6], so we added the instructions as text labels. To further attract users' attention, we added pulsing effect (where labels enlarged and reduced continuously) [12]. The system presents visual cues in multiple stages for interactive assistance (see Fig. 4).

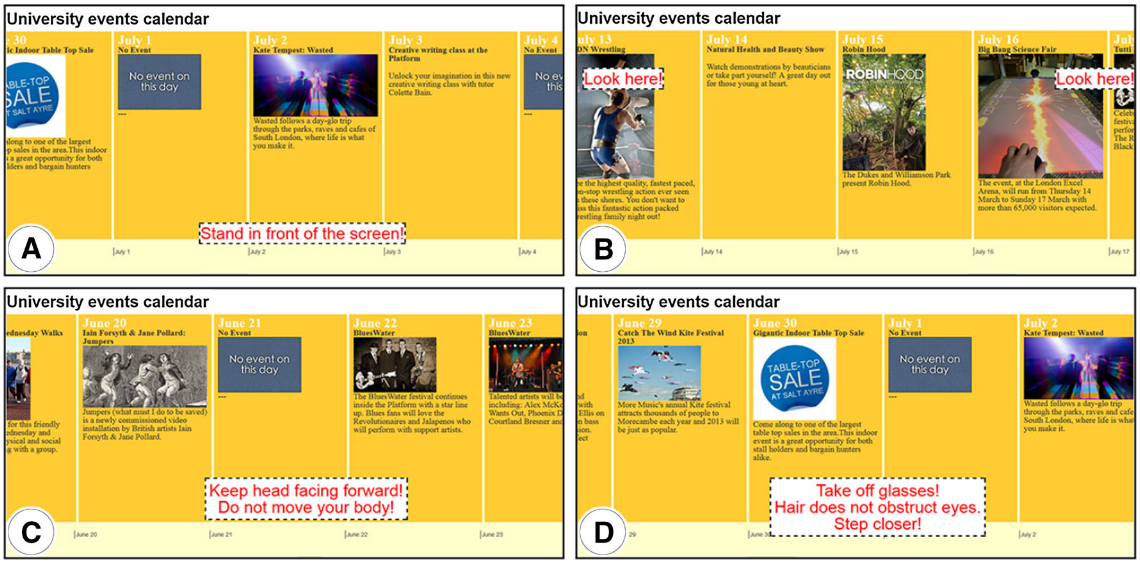
6.1 Set-up and method

We conducted a second field study to test whether users can translate the visual cues into user operation. We deployed our system in the reception area of a university research building. We used a 55-inch LED display, positioned at a height of 170 cm above ground, and a web camera was placed on the top edge, in the middle, of the display. We only invited novice users who did not participate in our earlier study. The conversations between the researchers and the participants were strictly limited to invitation, and we provided no assistance. Two members of our research team conducted the study, one for inviting passers-by and the other for interviewing participants with an exit survey.

6.2 Results

We conducted the field study over a 2-day period. In total, 35 passers-by (aged between 18 to 41, plus a child) tested our interface; 6 failed to use the system due to: strabismus (crossed eyes), myopia (blurred vision without corrective lenses), wearing tinted glasses, standing too close to the screen, not noticing the visual cues and a child whose height was too short. Interviews revealed that most of the participants found the visual cues informative—especially the "*Look Here*" label, but suggested that it was only needed for a short duration—and helped them to realise the interaction very quickly. They commented that the instructions were "*clear*" and "*self-explanatory*". Two users mentioned that the "Hair should not obstruct eyes" label was helpful for people with long hair fringe. Also, the pulsing effect naturally drew their attention, and it was very efficient for communicating interactivity.

We found that the majority of the participants followed the displayed instruction correctly. In general, the first two levels



Stand in front of the screen!

Look here!

Look here!

Keep head facing forward!
Do not move your body!

Take off glasses!
Hair does not obstruct eyes.
Step closer!

Fig. 4 (Field study 2) Stages of visual cues. **a** *Position*: In the first stage, the interface displays a message to invite users to stand in front of the display. **b** *Eye input*: Once the system detects the presence of a person, it displays "look here" labels, which indicates at where the user should look. *Corrective guidance*: **c** If the system detects the user is not facing forward (e.g. the head is turned), the system displays a "keep head facing forward" message. Also, **d** if the system detects the user's face, but not the eyes, the system assumes that something is occluding the eyes or the user is too far. The interface suggests the user to ensure their eyes are not occluded and to step closer when the detected face is too small

of information (see Fig. 4a, b) were crucial for apprehending the correct interaction. If no errors occurred, the visual cues for correcting guidance (see Fig. 4c, d) did not appear. A few users commented that some of the textual messages were too long and suggested further improvement could include graphical guidance (e.g. pictures) or simplified textual information (e.g. shortening phrases or highlighting keywords).

6.2.1 Floor marker versus on-screen distance information

As described in an earlier section, the distance of how far a user stands from the camera affects the tracking of the user's eyes. During our pilot test, the initial interface gave no precise information of how far and where the users should stand (see Fig. 4a). We noticed that our pilot users often stood too far (over two metres) away from the screen, so the system failed to detect their presence and remained non-interactive. This never occurred in the previous study as the set-up was different and a researcher gave assistance. To help users positioning themselves, this study also tested two approaches (see Fig. 5): (a) using a *floor marker*, by placing a "*Stand Here*" sign on the floor; **b** providing an *explicit distance information* on-screen, where the display showed a label informing users to stand at a distance of one metre away.

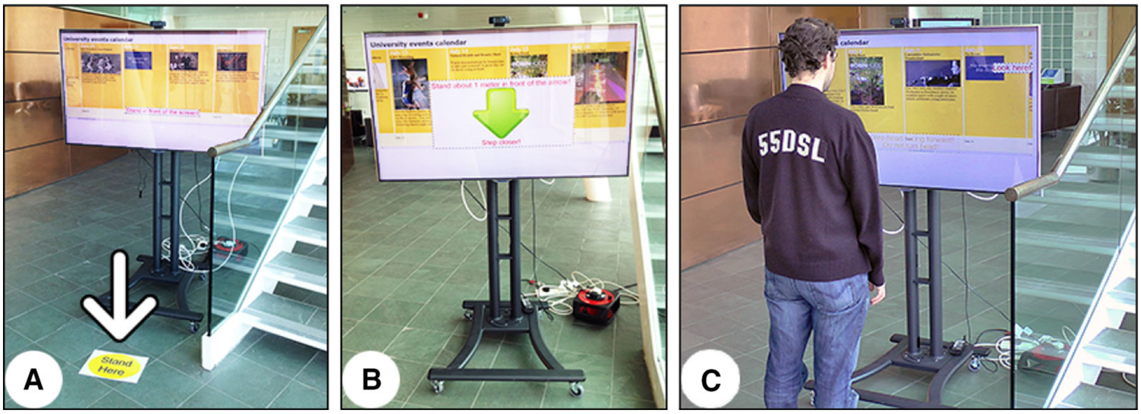


Fig. 5 (Field study 2) User positioning. **a** During the first day, we placed a foot marker on the floor, at a distance of 1 m from the screen. **b** On the second day, we removed the marker, and the interface displayed a *big green arrow* with a message saying "Stand about 1 m in front of the *arrow*". **c** A user testing our interface

During the first day of the study, we used a floor marker for helping users to position themselves. Some participants initially ignored the floor marker and only realised it later when they looked down. This indicates that people easily noticed the visual cues on the display, but not the cues on the floor level. On the second day, we removed the floor marker and the interface explicitly displayed the distance. This was more effective. All the participants noticed the cue, but required longer time for adjusting themselves to the correct distance and position.

For our next study, we decided to use an on-screen visual cue for assisting users to position themselves. It further has the benefit that the system is pure software based, which makes deployment simpler.

6.3 Summary

This study confirmed our translation of the minimum required levels from Field study 1 to interactive visual cues on the user interface. Our study showed that many people were able to follow a sequence of guidance labels on the display to figure out the interaction of GazeHorizon. Furthermore, we also learned that all visual guidance should be shown on the display (on the same level as the user's field of vision), otherwise labels placed outside the display could be unnoticed.

7 Field study 3: GazeHorizon in the wild

To understand how people unaidedly use our system, the objective of this study is to determine the general effects of GazeHorizon on passers-by in an ecologically valid setting. To maintain validity, we neither invited, interfered nor advised the passers-by; instead, participants were enticed purely by the interface.

7.1 Deployment, set-up and data collection

We implemented GazeHorizon as a browser of latest movies, and we deployed the system in the lobby of a university building, in Germany. Many people passed through this area everyday. They were mainly university students, staff and visitors. We used a 45-inch display, positioned at a height of 170 cm above ground, and we mounted a web camera on top of the display. To assist users positioning themselves, we added a mirrored video feed, overlaid with a face outline, on the interface (Fig. 6). Using video feed helped to communicate interactivity of the display [13].

During deployment, the system logged anonymous data of users' eye images and timestamped system events. We placed a video recorder opposite to the screen for capturing user behaviours. After the users finished their interaction, a

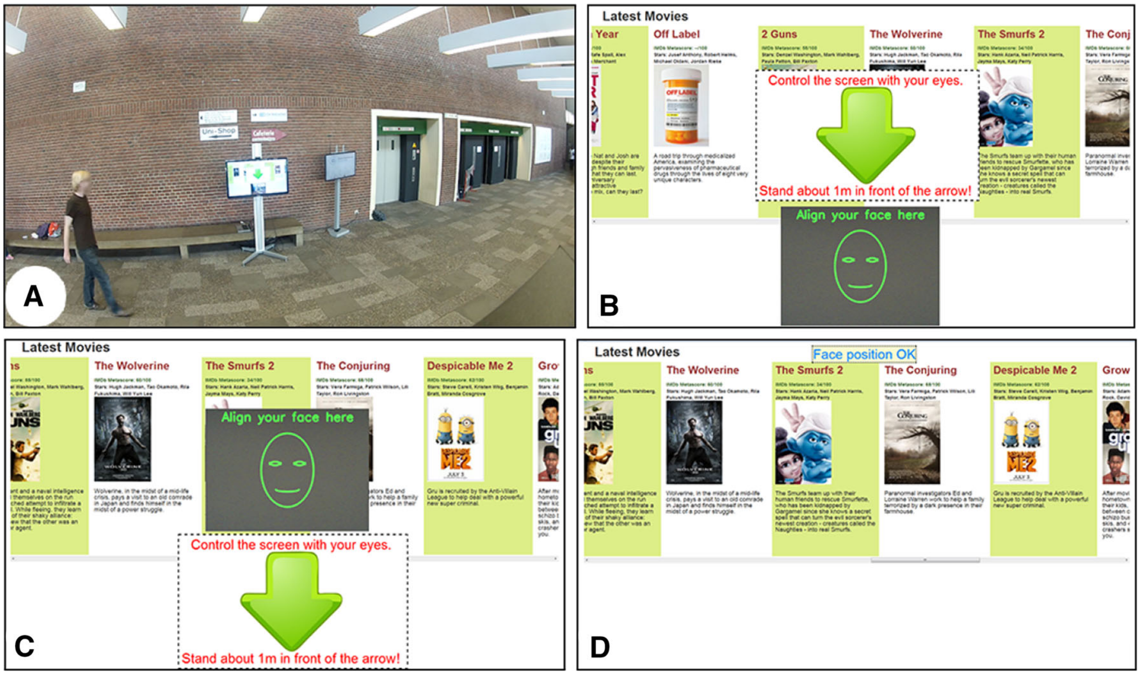


Fig. 6 (Field study 3) In-the-wild deployment. **a** GazeHorizon was deployed in a lobby of a university building. **b** Initial interface before optimisation. A large green arrow was used to attract user attention, and a mirrored video feed was displayed to help users to position themselves. Users often looked down to view the mirror image, which slowed down the detection. **c** After initial optimisation, we moved the mirrored video feed to the central region. **d** We also amended the interface by adding a "Face position OK" label for constant face alignment feedback

member of the research team approached the users for feedback. In the post-study analysis, two researchers independently analysed the log data, the recorded videos and the interview recordings.

We adopted Schmidt et al.'s two-phase deployment approach [17]: *Optimisation* and *Testing*. During days 1 and 2 (optimisation), we performed several iterations of improving the interface according to our observations and users' feedback. During days 3 and 4 (testing), the prototype was not modified, and we conducted detailed interviews with users.

7.2 Interface optimisation

During the first 2 days, we interviewed 46 users for qualitative feedback. We asked the users of the issues that they encountered and for suggestions for improvement. Based on the issues reported by the users and those we observed, we made amendments to the interface. This was done iteratively until a sufficient number of users were able to use the system.

In our initial interface, the mirrored video feed was positioned at the bottom of the screen (see Fig. 6b). When the users looked down, their eye lashes/lids often occluded their eyes, which prevented the system from detecting their pupil centres and eye corners. This inherently slowed down the system detection. We resolved this by moving the video feed to the centre of the display (see Fig. 6c).

Also, the original video feed was constantly shown at the bottom of the screen. Some users criticised that it was distracting. We changed the video feed to disappear after a user's face was aligned correctly. However, without the video, the users reported that they were unsure of whether their face was still aligned correctly. To provide constant feedback, we added a "Face position OK" label on the top region of the screen (see Fig. 6d). This label only disappeared if the user's face was out of alignment. Other minor changes include changing the colour of labels to give higher contrast.

7.3 Findings

The log data revealed a total of 129 interaction instances, where each instance contains a full episode of uninterrupted use, either by one or by more users [15]. Of the instances, 107 triggered continuous scrolling, with a mean interaction time of 67.1 s (SD = 54.2 s). Figure 7 shows a histogram of the interaction time. Most users interacted with our system for between 20 and 80 s. We were surprised that one user spent over five minutes. She explained that she really enjoyed movies, and she spent most of the interaction time reading the synopsis. From the moment when the system detected users' presence, on average, the users required 4.8 s (SD = 8.5 s) to align their face into the correct position and 7.2 s (SD = 11.0 s) to perform a scroll (also measured from the same beginning moment). Over the entire interaction duration, the users spent 27.0 % (SD = 15.1%) of the time for scrolling the content.

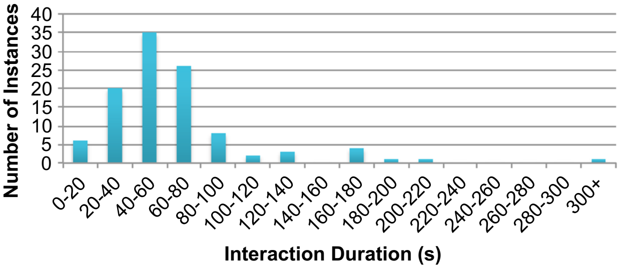


Fig. 7 (Field study 3) A histogram of overall users' interaction time over 4 days of deployment

During the optimisation phase, we interviewed 46 users, and 35 of them (76.0 %) reported that they were able to use the system for scrolling information. This rate increased after optimisation. We interviewed 41 users during the testing phase, and 35 of them (85.4 %) reported that they were able to scroll the content. Over the 4-day period, we observed 20 users who wore glasses, and 9 of them were still able to use our system without removing their glasses.

7.3.1 *Group behaviour and sharing experience*

Passers-by sometimes approached our system in groups, but usually one person interacted with our system at a time. People were more willing to try if they saw another person successfully used the system. If a user was able to comprehend the interaction, the user would encourage other members to experience the system, so the group were more likely to try. Also, people in a group helped others by pointing to or reading out the displayed instructions.

We observed the *honeypot effect* [2]. Passers-by became curious after noticing someone using GazeHorizon (see Fig. 8). Spectators first positioned themselves behind the user and observed from a distance, without disturbing the user. When the user noticed people were observing, the user often explained the interaction and invited the observers to try. We noticed instances where strangers engaged in short conversations to discuss about the operation of our system.

7.3.2 *Interacting with GazeHorizon display*

The textual label "Control the screen with your eyes" gave an obvious hint to users that the system is eye-based

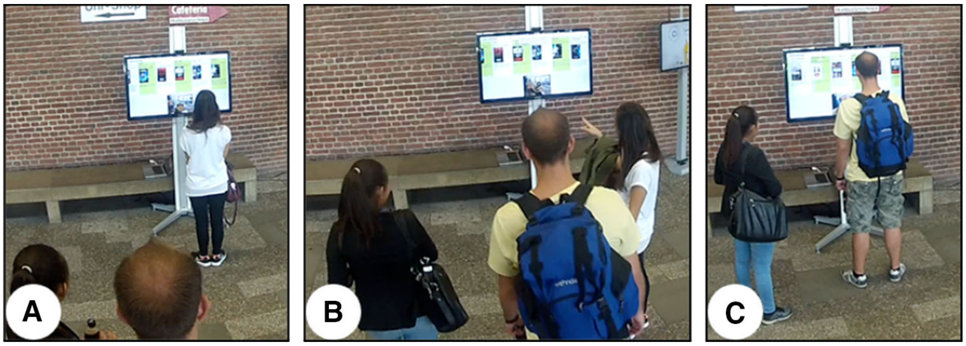


Fig. 8 (Field study 3) Honeypot effect. **a** Two passers-by observed a user. **b** The user explained the interaction. **c** The passers-by tried the system

interactive. The majority of users realised the interaction by noticing the movement of content when they looked at the display. Several people explained that the movie images first attracted their attention, and then, they realised the interaction model when the pictures moved. For example, "*followed it [a movie's picture] until the article [the synopsis] got slower ... in the centre it stopped*". Some users commented that they got comfortable with the system very quickly, and after 2–3 scrolling attempts, they realised that they did not need to stare at the "look here" label for scrolling content.

A few users explained that they were attracted by the novelty effect of "*using eyes to control*" scrolling. They were mainly interested in experiencing the new form of interaction and attempted the system to seek for different effects, for example, "*it scrolls faster when look more to the sides*". Other users who were interested in movies usually browsed through the entire content and interacted with the system for much longer.

Currently, gaze is still an uncommon modality, and many people are unfamiliar with using their eyes for interaction. When our users first approached the system, they sometimes did not read all the displayed cues, so they did not immediately know that the system was controlled by eye movement. Without knowing it was eye based, some people waved their hands to attempt to interact with the system (see Fig. 9a, b). After the users noticed no responses, they would then read the displayed cues and follow the given instructions. However, some users were impatient and abandoned further attempts after seeing no responses.

Although our system was designed only for scrolling content, some people expected the content to be selectable. We noticed that, after some users had successfully scrolled the content, they touched the screen in an attempt to trigger a selection (see Fig. 9c). Interviewees suggested using double touch, head nodding, facial expression (e.g. open mouth), blinks or winks, as well as stare at an object for a few seconds (i.e. dwell time), to activate a selection. Also, for some users, even though they knew the interaction was eye based, they attempted to use touch and swipe gesture to control the content (see Fig. 9d). Two users suggested "*vertical scrolling*" for more text when the content stops in the centre.

7.3.3 Common causes of failure

We analysed instances where users failed to scroll. We noticed that many instances were caused by people standing at the wrong location (see Fig. 10a, b). For instance, some users stood too far away from the camera; some noticed the video feed but did not want to be recorded, so they stood out of the camera focus. Either way, the camera failed to detect the users' presence. Another cause of failure was that, instead moving eyes, users turned their head towards the sides (see Fig. 10c). Although the interface indicated that users should keep their head facing forward, people sometimes missed that

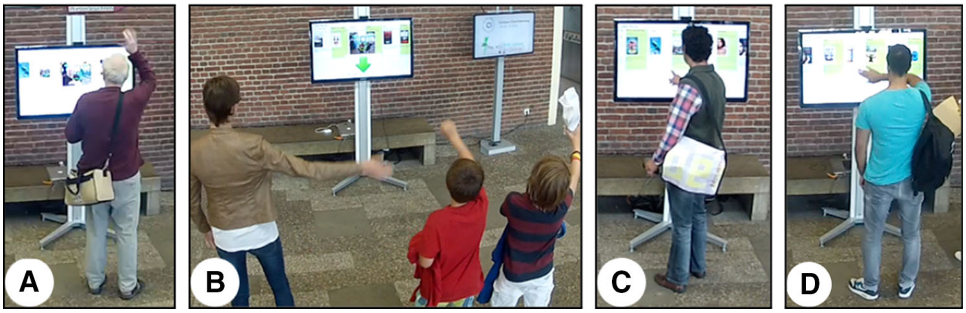


Fig. 9 (Field study 3) Users' expected interactions. **a, b** Users attempted to interact with the system by waving their hands. **c** A user attempted to make a selection by touching the screen. **d** A user attempted to scroll content by performing a touch and swipe gesture

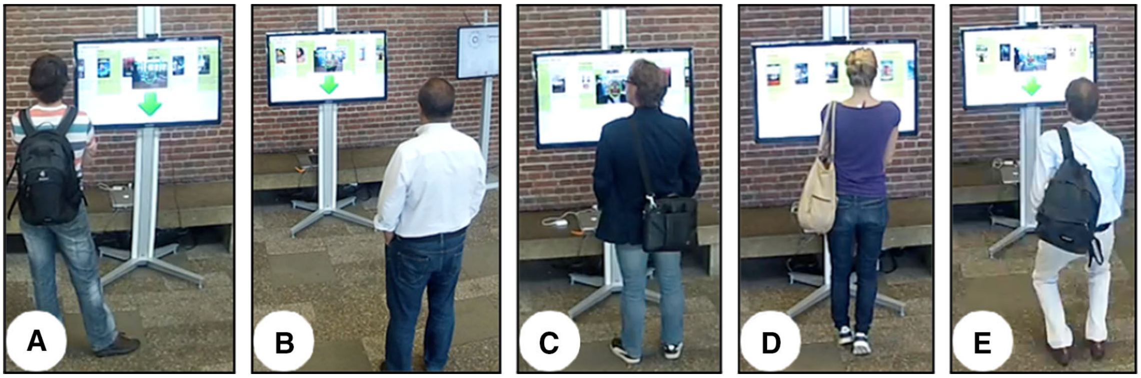


Fig. 10 (Field study 3) Examples of users' patterns of interaction. *Common causes of failure:* **a** standing on the side of the display. **b** Standing too far from the display. **c** Turned head to look at the side. *Height adjustment:* **d** a short user lifted her feet. **e** A tall user bent his knees

guidance label. Other common causes included: missed displayed labels, impatience due to system delays, could not see without glasses, and failure detection of user's face/eyes.

7.3.4 Self positioning

When novice users approached our system, they intuitively realised that they needed to align their face to the outline of the video feed. Although a label explicitly informed the user to stand one metre away from the display, interviews revealed that users had no reference for estimating the one-metre length. Instead, they judged their position by aligning their face. The users explained that they used the video feed for reference, as it provided real-time feedback.

When users positioned themselves, they were often able to stand correctly in the middle but too far away from the display. People stepped back and forth to adjust their distance and then fine-tuned by leaning/tilting their body, without moving their feet. Tall users tended to bend their upper body or their knees to lower their height, while shorter users lifted their heels (see Fig. 10d, e). We observed an instance where children jumped up and down to align their faces. To accommodate different height, the system can use a camera with a wider vertical angle for detecting users.

7.4 Users' feedback

We generally received positive feedback such as "*very promising*", "*useful*","*good for when hands are dirty and busy*" and "*great for the disabled*". The majority of users felt the system was "*really easy*", "*very fast to get used to how it works*", and the instructions were "*clear and helpful*". Some users commented that the scrolling interaction was "*logical*". They felt the system managed to "*captured [their] attention*" and the content "*changed with [their] view*".

Some users pointed out that the system was "*a bit slow with glasses*", "*works better when glasses were removed, but not effective as not able to read*". This was expected, as the shape of glasses frame can affect the face and eye detection. Other users also mentioned "*need to be patient*", "*takes too much time to make it [the content] move to the centre*". Delays varied between persons and also depend on lighting condition. The system works best in a bright environment.

7.4.1 Privacy concerns

In the testing phase, we prompted the users about privacy concerns while using GazeHorizon. The majority (34/41) reported that they were comfortable with using their eyes for scrolling information in public and did not perceive any privacy risks. Among those who were concerned (4 reported "Yes" and 3 "uncertain"), they noticed the web camera and they were worried about how the captured images were stored and used. They explained that it is acceptable if the data and their identity were not revealed publicly, but they preferred to have an option for opting out from being recorded. Also, the displayed content can impact their sense of privacy. One person particularly mentioned that information about movies was acceptable; however, other types of content (added with gaze information) may reveal personal interests unwantedly.

7.5 Lessons learned from GazeHorizon deployment

From the deployment, we confirmed that by providing intuitive guidance novice users were able to control a GazeHorizon display without prior training, and we learned that:

- Some users who could not use the system mentioned that they did not notice that the system reacts to gaze. Letting users know that the system reacts to eye movement at first glance is crucial. Knowing this up front helps users to eliminate attempts of other modalities, such as touch or gestures, and makes it easier for users to interpret the control of GazeHorizon during their first attempt.
- We observed that users lose patience very quickly. Some people abandoned further attempts if they see no immediate system response from their first action, and this is similar to the findings reported by Marshall et al. [10].
- The mirror image was more effective than text labels to assist users positioning themselves, as it provides real-time reference for users to perceive their position.

8 Discussion

Deploying eye tracking systems in the wild has long been a great challenge due to tedious set-up and adaptation to individual users. Our work makes three contributions to overcome this challenge. (1) GazeHorizon is a stand-alone vision-based software solution that only requires a single web camera mounted on the display. This simplifies deployment, as it requires minimum hardware set-up. (2) We present a robust person-independent method for supporting gaze interaction. Our field studies tested with over a hundred users and many of them could use our system. This demonstrates its real-world practicality. (3) Whereas conventional eye trackers mainly detect eyes, our vision-based system also detects other useful information based on the users' actions. For example, GazeHorizon tracks whether a user is approaching our system, and whether the user's head is turned. Our system interprets this context information to present dynamic guidance to assist the user in real time. Overall, we believe that vision-based gaze interactive system has a great potential for wide adoption as long as it achieves robustness in the real world.

8.1 Relative gaze mapping for rate-controlled navigation

While conventional eye tracking methods map gaze to absolute screen locations, we employed a relative mapping approach that provides different interaction experiences. Although relative mapping does not detect where on the screen users look at, our users' feedback revealed that they felt the system captured their view. This confirms our design that relative mapping can provide a robust illusion of display response to what the user looks at.

In absolute mapping, the reference is device centric to the screen space. Any error in estimated gaze direction will affect the user experience, as the display response will be relative to a screen position that differs from what the user actually looks at. In contrast, a relative mapping as adopted in GazeHorizon provides a user experience that is robust to inaccuracies in gaze estimation. An error in the estimate effects the scrolling speed but the user's illusion of content-of-interest moving to the centre of the display is robustly maintained.

8.2 Bootstrapping gaze interaction with interactive guidance

Eye movements are subtle. From our observations during the studies, users cannot learn gaze-based interaction by purely observing other users; instead, the learning process requires guidance. The guidance could be provided by either an experienced user explaining the interaction or interface guidance on the display. An experienced user could provide direct feedback and explanations; however, this relies on the experienced user understanding the interaction correctly. An alternative is via interactive guidance. We avoided to add explicit instructions; instead, we provided guided assistance when the system detects an anomaly. We believe that this is more effective and potentially reduces the cognitive load of users, as they discover the interaction model by exploring the interface at their own pace, and the guided assistance can help to prevent misconceptions and to correct user errors.

We learned that the "look here" label naturally captured users' attention. Although the intention of the users was primarily to look at the label, the action activated scrolling as an after-effect with no extra cost. From a novice user's perspective, the scrolling can be seen as an effect of his eye movement, which helps the user to conceptualise the activation of scrolling. We believe that the initial user experience was rather implicit; however, the interaction may become more explicit once the user understands the interaction. A few interviewees explained that once they learned the interaction, they explicitly moved their eyes to the sides for scrolling. Even though our interface did not provide any guidance for stopping the scrolling, somehow all of our participants self-discovered this operation.

From our field studies, we realised that there are many unpredictable factors that could hinder the tracking of users' eyes, such as unpredictable user behaviours. Causes of failure were often due to users standing too far away, in an incorrect position, wearing glasses or their eyes were occluded by their hair. They could be corrected by giving appropriate interactive guidance based on specific aspects. We realised that if users are aware of a particular reason that causes the system to stop tracking their eyes, the users are generally cooperative and willing to adjust themselves, like removing glasses, stepping closer. However, we observed an interesting behaviour that sometimes after users noticed the display, they would step away or stand on one side of the display to observe for a period of time. We consider this behaviour as *mirror image avoidance*: although users were curious to experience the system, they might deliberately position themselves to avoid being shown on the "mirror image". In some cases, users even moved around while kept looking at the display. This could be due to the users not knowing that the mirror image will disappear and they did not want to be recorded.

8.3 Group interaction

We implemented GazeHorizon as a single-user application. In our in-the-wild deployment, we observed that users in groups took turns to interact with the display individually. Nonetheless, applications in public spaces pose an issue of sharing an interaction space among multiple users. GazeHorizon could overcome this by distinguishing individual users from their faces and pairs of eyes, and the screen could be divided into multiple regions to support simultaneous interaction of multiple users in parallel. However, this inherently prohibits collaborations where users share the same screen space. A great challenge is thus to design interaction for group collaboration and to minimise confusions and conflicts between users. For example, a user might mistake that an action on the display was triggered by his gaze input, while the action was in fact triggered by another user. This leads to several open questions: How can we design interactions that give users a better perception of gaze control among multiple users? Also, when multiple users are involved, they are not aware of each other's eye movement; how can we design interface that promotes eye-based group collaboration?

8.4 Limitations

Our system only supports horizontal scrolling. We believe that further advances in computer vision could improve the expressiveness of GazeHorizon and extend the system to track vertical eye movement. This would allow scrolling of 2D content, e.g. maps and high-resolution images. We also envision that eye movements and facial expressions (e.g. emotions) can be combined to support richer user interaction.

Although we learned that placing visual stimulus can attract users' gaze attention, this could lead users to move their eyes, to turn their head, or both. Some users are not aware that the system reacts to eye movement initially, so they naturally turn their head to face towards the direction of their interest. Our system tolerates a small degree (30°) of head turning. If the head turned a small angle, screen objects still scroll but slower because the PCR is reduced. For large head turning, scrolling will not be triggered as the system cannot detect the users' eyes. Our results show that by providing appropriate interactive guidance, people were able to understand and adjust themselves to accommodate the limitation.

Our work is the first to demonstrate that gaze input can be used in public settings. Our studies show that novice users can easily apprehend the interaction of GazeHorizon. However, we have only explored the use of gaze; future work can compare or combine gaze with different types of input for public displays, such as combining gaze with head orientation.

9 Conclusion

In this paper, we presented GazeHorizon, a vision-based system that enables spontaneous gaze-based interaction on public displays. It employs a single camera and computer vision for person-independent tracking of horizontal gaze direction. We mapped this to rate-controlled navigation of horizontally arranged content. We conducted a succession of field studies and observed over 190 users (interviewed over 150 of them) to understand what guidance people require to discover the interaction of GazeHorizon. We evolved the system to provide visual cues and interactive assistance to bootstrap gaze interaction with unaware users. Finally, we deployed GazeHorizon "in the wild", where we neither invited nor assisted passers-by. The results showed that a large number of novice users successfully used GazeHorizon and were able to comprehend the interaction unassisted. Our work shows that it is possible to integrate spontaneous gaze-based interaction in public settings, and we believe that the work provides a foundation for the investigation of eye-based technology for public displays. We envision that gaze interactivity will enhance people's experience of acquiring public information, for example visitors viewing panoramic pictures in photographic exhibitions or scrolling a timeline in history museums, as well as customers browsing catalogues in retail shops.

Acknowledgments This work was supported by the EU Marie Curie Network iCareNet under grant number 264738.

References

1. Boring S, Baur D, Butz A, Gustafson S, Baudisch P (2010) Touch projector: mobile interaction through video. In: Proceedings of the CHI 2010, ACM Press, 2287–2296
2. Brignull H, Rogers Y (2003) Enticing people to interact with large public displays in public spaces. In: Proceedings of the INTERACT 2003, IOS Press, 17–24

3. Eaddy M, Blasko G, Babcock J, Feiner S (2004) My own private kiosk: privacy-preserving public displays. In: Proceedings of the ISWC 2004, IEEE computer society, 132–135
4. Hansen D, Ji Q (2010) In the eye of the beholder: a survey of models for eyes and gaze. IEEE Trans Pattern Anal Mach Intell 32(3):478–500
5. Jacob RJK (1991) The use of eye movements in human-computer interaction techniques: what you look at is what you get. ACM Trans Inf Syst 9(2):152–169
6. Kukka H, Oja H, Kostakos V, Gonçalves J, Ojala T (2013) What makes you click: exploring visual signals to entice interaction on public displays. In: Proceedings of the CHI 2013, ACM Press, 1699–1708
7. Kumar M, Winograd T (2007) Gaze-enhanced scrolling techniques. In: Proceedings of the UIST 2007, ACM Press, 213–216
8. MacKenzie IS, Zhang X (2008) Eye typing using word and letter prediction and a fixation algorithm. In: Proceedings of the ETRA 2008, ACM Press, 55–58
9. Mardanbegi D, Hansen DW, Pederson T (2012) Eye-based head gestures. In: Proceedings of the ETRA 2012, ACM Press, 139–146
10. Marshall P, Morris R, Rogers Y, Kreitmayer S, Davies M (2011) Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In: Proceedings of the CHI 2011, ACM, 3033–3042
11. Morimoto CH, Mimica MRM (2005) Eye gaze tracking techniques for interactive applications. Comput Vis Image Underst 98(1):4–24
12. Müller J, Alt F, Michelis D, Schmidt A (2010) Requirements and design space for interactive public displays. In: Proceedings of the MM 2010, ACM Press, 1285–1294
13. Müller J, Walter R, Bailly G, Nischt M, Alt F (2012) Looking glass: a field study on noticing interactivity of a shop window. In: Proceedings of the CHI 2012, ACM Press, 297–306
14. Nakanishi Y, Fujii T, Kiatjima K, Sato Y, Koike H (2002) Vision-based face tracking system for large displays. In: Proceedings of the UbiComp 2002, Springer, 152–159
15. Peltonen P, Kurvinen E, Salovaara A, Jacucci G, Ilmonen T, Evans J, Oulasvirta A, Saarikko P (2008) It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In: Proceedings of the CHI 2008, ACM Press, 1285–1294
16. Ren G, Li C, O'Neill E, Willis P (2013) 3d freehand gestural navigation for interactive public displays. IEEE Comput Graph Appl 33(2):47–55
17. Schmidt C, Müller J, Bailly G (2013) Screenfinity: extending the perception area of content on very large public displays. In: Proceedings of the CHI 2013, ACM Press, 1719–1728
18. Sippl A, Holzmann C, Zachhuber D, Ferscha A (2010) Real-time gaze tracking for public displays. In: Proceedings of the AmI 2010, Springer, 167–176
19. Smith BA, Yin Q, Feiner SK, Nayar SK (2013) Gaze locking: passive eye contact detection for human-object interaction. In: Proceedings UIST 2013, ACM Press, 271–280
20. Smith JD, Vertegaal R, Sohn C (2005) Viewpointer: lightweight calibration-free eye tracking for ubiquitous handsfree deixis. In: Proceedings of the UIST 2005, ACM Press, 53–61
21. Turner J, Alexander J, Bulling A, Schmidt D, Gellersen H (2013) Eye pull, eye push: moving objects between large screens and personal devices with gaze & touch. In: Proceedings of the INTERACT 2013, 170–186
22. Turner J, Bulling A, Alexander J, Gellersen H (2014) Cross-device gaze-supported point-to-point content transfer. In: Proceedings of the ETRA 2014, ACM Press, 19–26
23. Vertegaal R, Mamuji A, Sohn C, Cheng D (2005) Media eyepliances: using eye tracking for remote control focus selection of appliances. In: CHI EA 2005, ACM Press, 1861–1864
24. Vidal M, Bulling A, Gellersen H (2013) Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In: Proceedings of the UbiComp 2013, ACM Press, 439–448
25. Vogel D, Balakrishnan R (2004) Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In: Proceedings of the UIST 2004, ACM, 137–146
26. Walter R, Bailly G, Müller J (2013) Strikeapose: revealing mid-air gestures on public displays. In: Proceedings of the CHI 2013, ACM Press, 841–850
27. Zhai S, Morimoto C, Ihde S (1999) Manual and gaze input cascaded (magic) pointing. In: Proceedings of the CHI 1999, ACM Press, 246–253
28. Zhang Y, Bulling A, Gellersen H (2013) SideWays: a gaze interface for spontaneous interaction with situated displays. In: Proceedings of the CHI 2013, ACM Press, 851–860
29. Zhang Y, Bulling A, Gellersen H (2014) PCR: a calibration-free method for tracking horizontal gaze direction using a single camera. In: Proceedings of the AVI 2014, ACM Press, 129–132
30. Zhang Y, Müller J, Chong MK, Bulling A, Gellersen H (2014) GazeHorizon: enabling passers-by to interact with public displays by gaze. In: Proceedings of the UbiComp 2014, ACM Press, 559–563
31. Zhu D, Gedeon T, Taylor K (2011) Moving to the centre: a gaze-driven remote camera control for teleoperation. Interact Comput 23(1):85–95