SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays

Yanxia Zhang
Lancaster University
Lancaster, United Kingdom
yazhang@lancaster.ac.uk

Andreas Bulling
Max Planck Institute for
Informatics
Saarbrücken, Germany
andreas.bulling@acm.org

Hans Gellersen
Lancaster University
Lancaster, United Kingdom
hwg@comp.lancs.ac.uk

ABSTRACT

Eye gaze is compelling for interaction with situated displays as we naturally use our eyes to engage with them. In this work we present *SideWays*, a novel person-independent eye gaze interface that supports spontaneous interaction with displays: users can just walk up to a display and immediately interact using their eyes, without any prior user calibration or training. Requiring only a single off-the-shelf camera and lightweight image processing, SideWays robustly detects whether users attend to the centre of the display or cast glances to the left or right. The system supports an interaction model in which attention to the central display is the default state, while "side-long glances" trigger input or actions. The robustness of the system and usability of the interaction model are validated in a study with 14 participants. Analysis of the participants' strategies in performing different tasks provides insights on gaze control strategies for design of SideWays applications.

Author Keywords

Eye-based Interaction; Eye Tracking; Spontaneous Interaction; Situated Display; Calibration-Free

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces—*Input devices and strategies*

INTRODUCTION

Eye gaze is a modality of interest for spontaneous interaction with displays, as we naturally use our eyes to engage with displays we encounter, and as other modalities are often not as readily available as our eyes. However, eye gaze is difficult to harness for spontaneous and pervasive use. A limitation that undermines the idea of spontaneous interaction is that eye trackers generally require a calibration phase before interaction can commence, every time a user steps up to a display. An additional problem is the dependence of existing methods on specialist hardware and illumination of the eye, which can be difficult to provide on pervasive scale.

Figure 1. *SideWays* enables spontaneous eye-based interaction with displays: users can just walk up to a display and immediately interact with it using their eyes, without any prior calibration or training. In this example, the user controls a cover flow with looks to the left and right.

In this paper, we present a novel eye gaze interface, designed for users to be able to just walk up to a display and casually assume control, using their eyes only. Our system, *SideWays*, requires only a single off-the-shelf camera and distinguishes three gaze directions (to the left, right and straight ahead) as input states. These input states are detected in a spontaneous, robust and person-independent manner. There is no training or calibration involved, and no adaptation to individual users. Any user can step up to the display, and the system will be able to respond immediately to their attention (see Figure 1).

The interaction model we have designed for SideWays assumes a central region of interest on the display to which users align their gaze (e.g. guided by visual design of the interface), and adjacent areas to the left and right that users can select with "sidelong glances". A look straight ahead is interpreted as attention to the centrally displayed content. In this state, the interface is kept stable, and the eyes do not trigger any action. In contrast, glances to the left or to the right are associated with input actions. In terms of application logic, these glances are like pressing (and holding) a button, but the user experience may be more subtle and fluid with interface designs that have such actions appear natural and implicit.

In developing SideWays, we contribute as follows. First, we present a novel image processing method for detection of eye gaze directions. The new idea underlying the method is to measure the distance between inner eye corner and eye centre for both eyes in order to robustly classify gaze directions. For this, we have developed a processing chain including a novel method for eye centre localisation (using colour edges as they

prove more robust under real world conditions). Secondly, we describe a study in which we evaluated SideWays with 14 participants on three interactive tasks (selection, scrolling, and slider control). The selection task served to characterise the system in terms of correct detection of input depending on time window of observation of the eyes, while the other two tasks assessed usability of the interface and interaction model for control tasks. Thirdly, we analyse the participants' behaviours in performing the different tasks. This lets us draw out insights on user's gaze control strategies, and derive design considerations for application of our system.

RELATED WORK

Eye-based Interaction
Early work on eye gaze interaction demonstrated selection on menu grids [6] and gaze-based techniques within the WIMP paradigm (windows, icons, menus, pointers) of our desktop computers [7]. The underlying model is to treat gaze as a pointing device, for example as alternative to mouse use for object selection in conventional interfaces [18], or for fast hands-free input with special-purpose visual interfaces [26]. In conventional settings, pointing accuracy is achieved with remote eye gaze trackers that require users to calibrate their gaze prior to interaction [10]. In contrast, we target pervasive settings and adopt a deliberately simpler gaze model to facilitate spontaneous and calibration-free interaction, between users and displays that have never seen each other before.

Use of the eyes for interaction with pervasive devices has previously been explored in work demonstrating attentive user interfaces that respond to eye contact [24]. This has been realised with pervasive eye contact sensors mounted on "smart" devices and objects [23], as well as with wearable devices that detect gaze orientation towards infrared tags placed in the environment [20]. In SideWays, we likewise detect gaze for attention to pervasive displays but in addition enable users to provide input with looks to the left or right from their centre of attention. Contrasting previous work, we do not require special-purpose hardware but detect eye gaze with a camera.

Sideways targets interaction with displays we encounter in our environment. A large body of work provides general insight into how people interact with displays in public settings (e.g., Opinionizer [1], Citywall [15] and Looking glass [12]), which is of importance to inform practical deployment of Sideways. A variety of projects have used head orientation towards large displays in presumed approximation of what people look at [11, 13, 19]. However, Mubin et al. found in an "interactive shop window" study that only few users aligned their heads with their gaze [11]. Other work has focused on low-cost extension of public displays for gaze pointing however still requires a calibration phase prior to interaction [16]. Contrasting extension of display, EyeGuide [3] explored use of a wearable eye tracking for interaction with pervasive displays. In contrast, our focus is on enabling interaction with public display without any instrumentation of the user. In our approach, all a user needs for interaction is their eyes.

Magee et al. reported a vision-based system that is similar to ours as it detects rapid eye movements to the left and right as command input [9]. However their system was specifically designed for a user with severe cerebral palsy and was primed to detect occurrences of a left/right movement, while we continually classify eye gaze direction. Kumar et al. presented several gaze-enhanced scrolling techniques for manual and automatic scrolling which use gaze as primary or complementary input [8]. In this work, scrolling is among three tasks on which we evaluate usability of our gaze interaction system.

Eye Tracking and Gaze Estimation
Research on gaze estimation mostly focuses on *model-based* approaches, such as Pupil Centre Corneal Reflection (PCCR) techniques [28, 29]. These techniques require specialised hardware and additional light sources, such as infrared illumination [5]. Gaze direction is estimated by detecting the pupil centre and a reflected "glint" on the cornea. With a geometric model of the eye, accurate gaze estimates are derived from the offset of the glint. However, the approach is problematic for pervasive settings as it requires controlled lighting conditions for illuminating the eye and for robust detection of light reflected from the spherical surface of the cornea.

Most commercially available eye trackers are based on model-based approaches and are typically geared toward desktop use. These systems require user-dependent calibration [2, 10]. More sophisticated eye trackers use stereoscopic views with multiple cameras or multiple light sources for multiple glints to avoid calibration for each individual [5]. A few user-independent eye-trackers using model-based approaches have been shown to work in laboratory conditions [28].

Recent efforts aim to avoid specialised hardware and infrared light sources to estimate gaze. *Appearance-based* approaches work under visible light and use image processing and machine learning techniques to estimate gaze directly from images recorded with video cameras [14, 21, 27]. Such approaches avoid calibration but require a priori training with eye images of the users, which prohibits spontaneous interaction. Sesma et al. conducted an offline evaluation of the pupil centre-eye corner vector as a feature for gaze estimation using a webcam-based eye tracking system [17]. However, their approach is person-dependent and initial calibration is required for each user to obtain the interpolation equation.

THE SIDEWAYS SYSTEM
The SideWays system is designed to detect gaze attention to a display, and to support gaze selection of areas to the left and to the right of the display's centre. The only hardware required for the system is a single off-the-shelf RGB video camera, positioned to capture the scene in front of the display. The system takes video frames as input, and analyses these with image processing techniques. Figure 2 illustrates the processing chain, from face detection and segmentation of eye images to detection of inner eye corners and localisation of pupil centres. The output from image processing are the distances of the eye centre from the eye corner for both left and right eye. Our system uses these distances to discriminate gaze directions as L(left), R(right) and C(centre). We explain both the image processing chain, and the eye gaze classification method in detail.
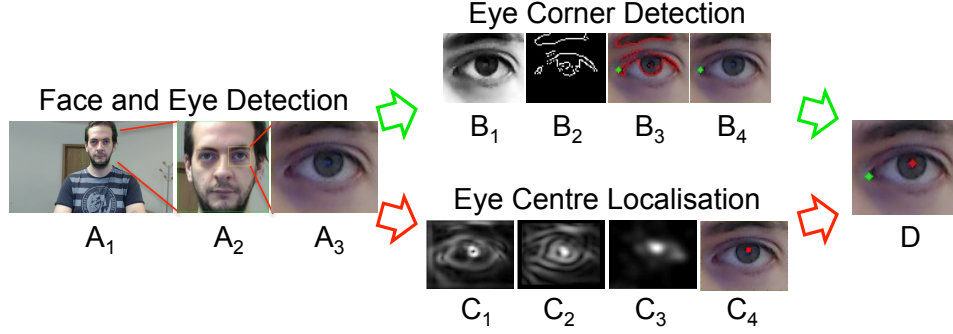
Eye Corner Detection

B₁ B₂ B₃ B₄

Face and Eye Detection

A₁ A₂ A₃

Eye Centre Localisation

C₁ C₂ C₃ C₄

D

Figure 2. This diagram illustrates the image processing of the SideWays system

Image Processing

Face and Eye Detection

The images from the video camera are subject to variable lighting conditions, such as shadows, bright lights, low contrast, motion blur or noise. To tackle these problems, we first apply a bilateral smoothing filter to reduce noise while preserving edge information. The filter smooths regions while preserving, and enhancing, the contrast at sharp intensity gradients. After the image is pre-processed, we use a Viola-Jones face detector to detect the user's face and eyes in real time [25] (see Figure 2: A_2). The face detector identifies a rectangular area of the largest face in the scene. In order to improve performance and reduce the computational cost, we only search for eyes in the top half region of the face. The system further detects the left eye by examining the left half of the facial area, and likewise for the right eye. Our system finally extracts two image patches that represent the output of this first processing step, one for each eye (A_3).

Detection of the Inner Eye Corners

To detect the inner eye corners, where the upper and lower eyelids meet (the so-called eye canthi), our system applies a Harris corner detection separately to each eye image patch. The method detects the locations of interesting windows which produce large variations when moved in any direction in an image. Locations which are above a threshold are marked as corners. Besides eye corners, the output could also include other feature points of local intensity maximum or minimum, line endings, or points on a curve where the curvature is locally maximal in the image. As this may result in several candidate eye corner points, the system further applies a canny edge detector (B_2). A canthus region should exhibit strong edges where the upper and lower eyelid edges meet. Accurate canthi locations can then be obtained by considering only those candidate points that lie on the detected eyelid edges (B_3). All other candidate points are discarded.

Localisation of Eye Centres

To localise the eye centres, we have develop a novel method that exploits the semi-circular structure of the eye iris and pupil as described in [22]. The captured eye images are strongly influenced by shadows cast in inner eye corners and by screen reflections. To reduce this effect, colour edges are obtained from cropped RGB eye images by using the Gaussian colour model proposed by [4] (C_1). The isophotes of an

image are curves connecting points of equal intensity [22]. We compute the isophote curvature map and calculate the isophote radius on the edges (C_2). The eye centre is detected as the location with maximum isophote centre votes (C_3)).

Discrimination of Gaze Directions

In each video frame, we obtain the pupil centre (C_4) and the inner canthi (B_4) as result of image processing. Figure 3 illustrates how we use these to derive gaze direction. Consider first that we look straight ahead. In this case the pupil centres of both our eyes will be similarly distant from the respective inner eye corners. If we look to the left, then the distance of our left pupil from its inner eye corner increases, while the distance of the right pupil from its inner corner decreases. Conversely, a look to the right means that the left pupil moves closer to its inner eye corner, while the right pupil moves further away. Consequently, to determine different gaze directions, we calculate the ratio r of the eye-centre P_{cx} to inner canthi C_{ix} distances of both eyes as

$$r = \left| \frac{C_{iR} - P_{cR}}{C_{iL} - P_{cL}} \right| \qquad (1)$$

where C_{iR} and C_{iL} are the x coordinates of inner eye corners from the right and left eye images, P_{cR} and P_{cL} are the x coordinates of pupil centres from the right and left eyes.

For each video frame, we calculate the inner canthi to pupil centre distances for both eyes. At time t ($t = 1, 2.., T$), we denote the measurement of our observed gaze direction for frame f_t as O_t. A general threshold T_r is set for classifying frame f_t as three gaze directions which are L(left), C(centre), R(right) according to the following rules:

$$O_t = \begin{cases} R(right), & \text{if } T_r < |r| < T_{r_{MAX}}; \\ L(left), & \text{if } 1/T_{r_{MAX}} < |r| < 1/T_r; \\ C(centre), & \text{otherwise.} \end{cases} \qquad (2)$$

where $T_r = 1.3$ is optimised for the user study (derived from the preliminary study) and $T_{r_{MAX}}$ is a constant upper bound.

Activation Using a Smoothing Window

To smooth the decision made for each video frame, we adopt a sliding window approach (as shown in Figure 4). We perform a smoothing window with a size of W over the observations of W image frames. We consider all observations within time span $[t, t + W - 1]$. The input to the decision system is
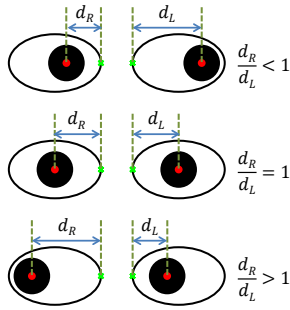
Figure 3. Three gaze directions are determined by the distance ratio of eye-centre P_{cx} to inner canthi C_{ix} for both eyes.
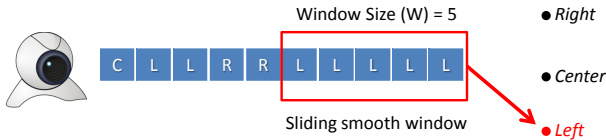

Figure 4. The activation of a gaze direction is based on a smoothing window. As the camera captures images in real-time, a stream of gaze directions fill the frames in the sliding window. An activation is triggered when the same gaze direction is detected consecutively.

a set of measurements $O_{1:W} = \{O_t | 1 \leq t \leq W\}$ where $O_t \in \{L, C, R\}$ in a sliding window of W frames within time span $[t, t + W - 1]$. We process all images frame by frame from the start. For each incoming frame, we collect new observations denoted as $O_{current}$. The set of measurements $O_{1:W}$ corresponding to time span $[t + 1, t + W]$ in the buffer is updated. We only consider valid observations when both eye corners and centres are detected. As shown in Figure 4, an activation is triggered when the same gaze direction is detected consecutively in the sliding window.

We distinguish two different activation approaches: *discrete* and *continuous*. Discrete activation clears all measurements $O_{1:W}$ in the sliding window after an event is triggered. This causes delay as a new stream of measurements in the sliding window need to be collected. Continuous activation only updates the last measurement in the sliding window with every incoming frame, hence, allowing for fast response.

USER STUDY

The interaction model proposed for the SideWays system is to treat gaze straight at the display's centre as a default state in which the eyes do not trigger any action, while "sideways" glances to the left or right are foreseen for user input. We designed a user study to evaluate our system and the proposed interaction model on three generic tasks: *Selecting*, *Scrolling* and *Sliding*. For each task, we run a separate experiment to evaluate different aspects of our system.

Selecting was always conducted first as it was designed to fundamentally characterise the interface in terms of correct classification of input depending on size of the sliding window used in the process. A smoothing window of five frames was used but data was collected for post-hoc analysis of detection accuracy versus speed (shorter smoothing windows). The other two experiments were conducted in counter-balance.
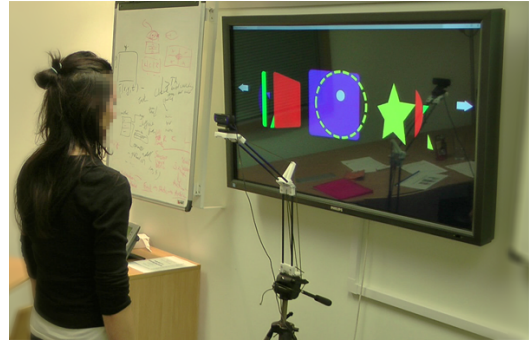

Figure 5. A snapshot of a study session.

Scrolling tested the users' ability to use our system for discrete scrolling through a list of items. A window size of four was used with discrete activation, so that a scroll step was executed only if the user's gaze dwelled for four valid frames on the left/right control. Sliding tested control of a continuous slider and the users' ability to move a slider to accurately hit a target position. A window size of three was chosen with continuous activation, which meant that that a sliding step was executed in each frame, for as long as the detected gaze direction matched the previous two frames. Three different speeds of the slider were used, and data captured to analyse how often users needed to change sliding direction to reach the target.

Participants
Fourteen paid participants (six female, eight male), with body heights ranging from 1.65m to 1.96m (*M*=1.77, *SD*=0.10), aged between 21 and 47 years (*M*=28.79, *SD*=7.27), and various eye colours took part in the study. Three participants wore contact lenses during the study.

Apparatus
The hardware setup for our study consisted of a 55 inch (121cm×68.5cm) LCD display from Philips with a resolution of 1920x1080 pixels, mounted on the wall at 120cm height (lower bezel). A Logitech HD Pro Webcam C920 with a resolution of 1280x720 pixels and a video frame rate of 30Hz was mounted on a stand and positioned 60cm in front of the screen (see Figure 5). The real-time image processing and gaze estimation software was implemented in OpenCV, and ran on a laptop with a 2.67GHz processor and 4GB of RAM.

The study was conducted in office space under normal lighting conditions. We asked participants to stand at a distance of 1.3m in front of the display (visual angle of the display 49.9° horizontal, 29.5° vertical). A marker on the floor indicated where the participant should stand. However, during the user study, participants were free to fine tune the distance for their own comfort. The distance between the camera and the user was 70cm±5cm. The captured image resolution was 300x300 for faces, and 80x70 for eye images, slightly varying across users. In a real world deployment, cameras would typically be mounted on the display but we positioned it closer to the user as we aimed to evaluate interaction with our system, not the limits of eye pupil and corner detection.

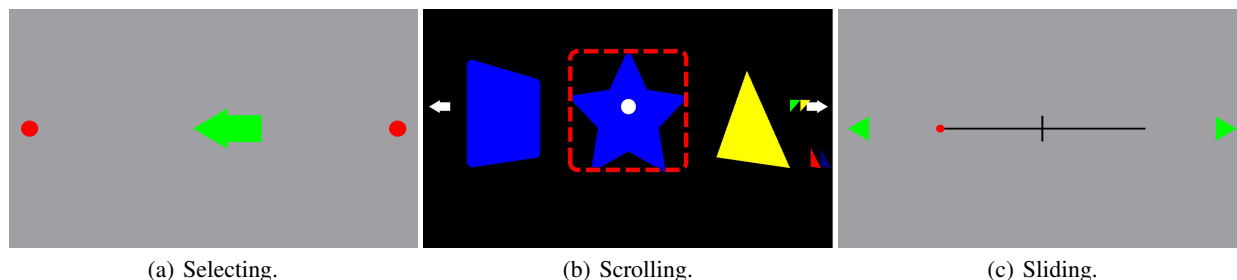| (a) Selecting. | (b) Scrolling. | (c) Sliding. |

Figure 6. The interfaces of the three tasks in our user study. (a) The participants read the green arrow to find out which direction to look at and stare at the corresponding red stimulus. (b) The participants scroll through a set of objects to find the one that matches a predefined target. (c) The participants control a horizontal slider and bring the slider indicator on top of the vertical bar.

Procedure
Each session lasted for approximately 45 minutes. Participants were first introduced to the system and allowed to complete one trial of each task. All participants then first completed the Selecting experiment, while the remaining two experiments were counter-balanced. After each experiment, user feedback was collected with a questionnaire. The questionnaire asked for the participants' subjective experience, problems they encountered, and strategies to overcome issues of the system.

Experiment 1: Selecting
In the selecting task, participants had to look at either the left or the right region of the display. Participants were asked to initially focus on the centre region of the display. Once the system had detected the participants' gaze, the system indicated the desired gaze region using a green arrow pointing either left or right. In addition, a red circle was shown in both the left and right region to assist participants in fixating on that region (see Figure 6(a)). The system continuously estimated the gaze direction with a smoothing window of five frames. Upon detection of gaze on the correct target, the target colour would change from red to green. Participant's were instructed to return their gaze to the centre after each completion of a trial. With a short delay, the next trial would be triggered by display of an arrow. In total, this was repeated twelve times (six for each direction, randomised).

Experiment 2: Scrolling
In the scrolling task, participants were asked to scroll through a list of objects using their gaze and to find the object that matched a predefined target. We used a combination of four shapes (circle, square, triangle, and star) and four colours (red, green, blue, and yellow) to represent a set of sixteen scrolling objects (see Figure 6(b)). At the beginning, the sixteen objects were randomly placed horizontally at equal distances, arranged as a flow with the display as viewport showing the current selection in the middle, and one adjacent object on either side. Participants then had to scroll through the objects to find a preselected object that was indicated by a coloured dash-bordered shape and shown at the centre of the display. Participants had to look left or right of the display to scroll items from that direction toward the centre. Arrows were displayed on both sides of the display to help participants fixate. The task was repeated six times, and each time with a dif-

ferent target shape. For each iteration, the browser starting position alternated between the left-most and the right-most positions of the object collection.

Experiment 3: Sliding
In the sliding task, the participant's objective was to control a horizontal slider by moving the slider indicator either left or right onto a target position with their eyes. For this task, the display showed a horizontal slider widget in the centre region (see Figure 6(c)). The slider target was represented by a black line, and it contained a red circle as slider indicator. Green arrows on the left and right display regions represented the slider's controls. At the start of the task, the indicator was placed at either the left-most or the right-most slider position, and the distance to the target was always 480 pixels. When participants looked at the left controller, the slider would progress one step width to the left in each frame, and vice versa for the right control. Sliding speed was increased over the trials, with three step widths representing 0.01, 0.025 and 0.04 of total screen width (i.e. 19.2, 48, and 76.8 pixels), requiring 25, 10 or 6 steps respectively to reach the target. When participants were satisfied that they had reached the target, they returned their gaze to the centre of the display to complete the trial. The task was repeated twice for each step size, for a total of six trials per participant.

RESULTS
Four of the 14 participants required eyeglasses for correct vision and three wore contact lenses while the fourth removed his glasses. One of them reported that the contact lenses had affected her speed in fast and frequent eye movements, and that they caused discomfort after using the system for a while.

Two participants experienced asymmetric system performance in left and right directions. One of them explained that she had had an eye operation, which effected gaze to the right. The other participant reported better performance for gaze to the right but her reason was unknown. However, she described that she compensated by turning her head slightly towards the opposite direction. The participant who removed his glasses was far-sighted. He often squinted his eyes while looking at the display, which drastically slowed down the system's detection speed (more frames were discarded for lack of pupil detection, and the smoothing window would fill up more slowly). Several participants reported that blinking also

	w = 2	w = 3	w = 4	w = 5
Correct	135	151	160	162
Wrong	12	5	2	0
Missed	21	11	6	6

Table 1. Number of correct, wrong (detected as opposite direction) and missed (detected as center) detections of the total 168 trials of selecting tasks using different window sizes.

	$t < 2$(s)	$t < 3$(s)	$t < 4$(s)	$t < 5$(s)	overall
w = 2	76.8%	78.6%	79.2%	80.4%	80.4%
w = 3	80.4%	86.9%	89.3%	89.9%	90.4%
w = 4	81.5%	87.5%	90.5%	91.1%	95.2%
w = 5	72.0%	84.5%	88.1%	91.1%	96.4%

Table 2. Accuracy for selection over different window sizes and timeout thresholds.

reduced the system's detection speed. This was not surprising given that blinking caused the eyelids to occlude the eye pupils, so the system could not determine their gaze direction.

Selecting Task

Participants performed a total of 168 trials (12 each) of looking left and right for selection. Table 1 summarises the results with post-hoc analysis for different window sizes. Eye gaze matching the target was counted as correct, eye gaze on the opposite target counted as wrong (faulty selection), and eye gaze in the centre as missed (no selection). Window size 5, as experienced by the users in the study, results in the least number of errors. This was expected, as increasing the window size also increases certainty. Analysis of recorded data revealed that three of the errors (missed selections) were from one participant due to squinting. On average, users needed 1.78s (*SD*=0.3s) per selection, but the user who squinted required an average of 2.91s. Although increasing the window size provides better detection accuracy, it inherently increases the time required to detect a correct selection. Table 2 provides an analysis of correct detection rates depending on window size and time thresholds. The results indicate that a window size of four frames is optimal, but windows of three frames perform almost as well for detection in limited time.

Participants reported that using our eye-based system for left and right selections was "intuitive", "easy to perform", and "suitable for touch-free interactions". However, two mentioned problems, such as inconsistency in system response time, one of them noting that to improve the system, he had tried not to blink and kept his eyes wide open. Several participants mentioned that although it wasn't tiring to use the system they would prefer faster response times.

Scrolling Task

We observed 84 trials for the scrolling task, six per participant. In each trial we counted how many scrolling steps users required to complete the trial and compared this with the minimally required steps. On average, participants required 1.2 (*SD*=3.4) extra steps to complete a scrolling trial. In 61 out
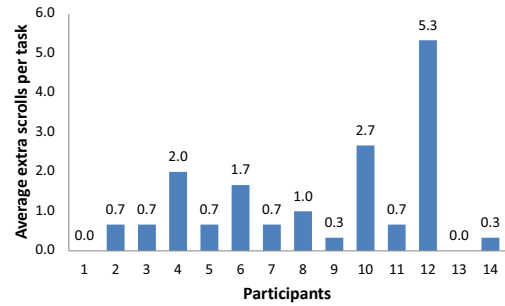

Figure 7. Extra scroll counts for 14 participants over six scrolling trials.

of the 84 (72.6%) trials the scrolling task was completed efficiently with the minimum number of steps. In general, participants were able to correct mistakes with very few extra steps. In 20 out of the 23 error cases, the participants only changed the scrolling direction once for correction (12 for overshooting one step away from the target, eight for a distance of two steps from the target).

Figure 7 illustrates the scrolling accuracy of each participant. Two participants finished every task without any extra scroll steps, and evidently learnt to use the system very quickly. Seven participants made one error (an extra scroll). However, participant 12 took 28 extra scrolls for his first trial, explaining that he lost focus during the first trial as the target shape was at the very end and the system was not responsive. Our post-study analysis showed that his head orientation drifted from the centre towards the scrolling direction, where new shapes were coming from. This caused the system to classify his gaze as central, instead of triggering a scroll step. The participant learnt to re-centre his head when the system was not responsive, and completed the remaining trials with only minor errors. Overall, we observed that six participants tended to turn their head towards the scrolling direction. Since the focus region is on the far end of the display, people intuitively turned their head towards the region to examine upcoming information. However, when they noticed the scrolling stopped, they returned their head orientation back to the centre.

System errors mainly resulted from delays in stopping the scrolling. Four participants mentioned that the system was not sensitive enough for stopping and that they had to look at the centre region already before the target object reached in the centre. In particular, participants 10 and 6 found it difficult to stop scrolling, while participant 6 found it hard to judge the colour of objects in the centre with peripheral vision, while gazing to the left or right for scrolling.

We observed the participants' behaviours and strategies in scrolling. Six participants fixated on the arrow indicator to scroll continuously. The participants were able to finish 40 out of 84 (47.6%) trials without stopping to scroll before reaching the target. Some participants mentioned that frequent stopping and checking the information in the centre helped them to perform better, but those who scrolled without stopping did not cause more errors (six errors in 40 trials). These results show that the participants were able to handle frequently changing information when the shapes are moving.

	Mean (s)	Std (s)	Minimum (s)
Large	14.2	12.0	2.4
Medium	16.2	11.5	6.0
Small	26.7	18.8	7.1

Table 3. Average and minimum completion time in seconds for three different step widths in the sliding task.
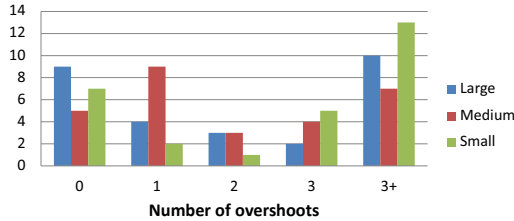

Figure 8. Histogram shows the count of the number of overshoots (0, 1, 2, 3 and > 3) for three different step widths in the sliding task.

The participants devised strategies to avoid mistakes. Some participants noticed the delay in triggering single scroll/stop actions (the system was set to discrete activation after four frames), and exploited this for brief glances to the centre without causing to stop the scrolling action.

Overall, participants were satisfied with their experience of using SideWays for scrolling. Most felt that the system reacted to their left and right gazes correctly, and that the system provided sufficient precision for real applications. They also felt that it was convenient to search objects using only their eyes, and suggested that the system is suitable for controlling "objects beyond reach". The participants enjoyed the experience of searching in a smooth flow, "without clicking". Most remarked it was easy and natural to use their peripheral vision for searching in this task. Given the big object size and simple content, they were able to see what was in the centre while looking at the scroll arrows. However, several participants mentioned that they needed to keep their head still, which they found difficult for longer scrolling. Exaggerated eye movements (e.g. changing from left to right) to correct mistakes caused fatigue. In addition, the participants preferred faster triggering time.

Sliding Task

This task tested the participants' ability to accurately move a slider to a target position. Participants performed six trials, two for each of three different slider speeds, and we collected data on a total of 84 trials, 28 per slider step widths. Table 3 shows the participants' average completion time. The last column (minimum) indicates the fastest time that the participants achieved. However, the average completion time was much higher. Most of the time was spent on position fine tuning. Many participants missed the slider target, thus requiring longer time to correct the slider's position.

We define overshooting as the number of instances when the indicator had jumped past the target location. Figure 8 provides a histogram that summarises observed overshooting. In 25% of all trials, participants managed to slide directly to the target without overshooting. Repeated overshoots indicate

problems with accurate control, and occurred more frequently with small and large step width. The mean of overshoots was 2.7 for medium step width, but 4.3 for large and 5.0 for small step width. Note that trials always started with small step width and that results will be influenced by a learning effect.

The errors caused by the fast speed (large step width) were mainly caused by system delay. Six participants criticised that the system was not fast enough to react when gaze direction changes rapidly. In addition, several participants also mentioned that the fast speed was too fast for their eyes. One participant was having difficulty with small step widths (12.5 overshoots/trial) while performing well with the medium and fast speed (1.5 overshoots for medium steps, 2.5 for large steps). The participant was short-sighted and removed her glasses during the study. She was not familiar with stopping with the initial trials, and followed the indicator moving left and right repeatedly crossing the target.

In general, the participants found it difficult to control the sliding indicator precisely with their eyes. Their strategy was to first bring the indicator near the target location as close as possible. This was done by staring at the control arrow for continuous sliding, while using their peripheral vision to approximate the indicator's location. Once the indicator was near the target, the participants looked at the centre region to stop the sliding. They fine tuned the movement by looking at the arrow control and the centre region back and forth. For the slow and medium speeds, in many trials, the participants were able to stop before the indicator reached the target (out of 28 trials, 18 times for slow and 16 times for medium); however, with faster speed they struggled more to do that (10 out of 28 trials). On the other hand, the slow speed caused issues in fine tuning. The participants reported that it was difficult to control with precision using peripheral vision. Since the distance of each jump was small, it was difficult to judge when exactly it reached the target. Some participants struggled to use their peripheral vision, and could not see both the arrows and the target together simultaneously.

Overall, most participants found the sliding task challenging when using the fast and the slow speeds. A few participants suggested the system could be useful for moving objects in out-of-reach distance. Several participants liked the fast and accurate response of the system. They found it easy to control the direction of the sliding object by using left and right eye movement. The majority felt that they were not able to control the system for fine tuning positions, especially using small step width. Half of the participants mentioned that it was unnatural to use their peripheral vision to see detail in the centre, while looking left or right. Some participants felt that they needed to concentrate and be patient to use SideWays for sliding. Also, the participants experienced fatigue due to frequent eye movements changing between left and right for position fine tuning. A few participants disliked the long time required to precisely move an object to the target.

Questionnaires

For each task, we asked seven questions with regards to the participants' subjective feedback. The results are presented in Figure 9. We run the Friedman Test on the subjective
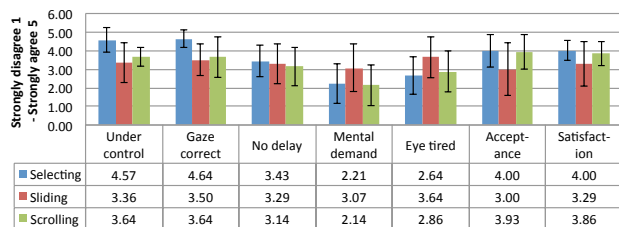
Figure 9. Participants' subjective feedback on use of Sideways for the selecting, scrolling and sliding tasks. Users were asked: did they feel in control; did the system respond correctly to their gaze; did the system respond without delay; did they find the task mentally demanding; did they find it tiring their eyes; would they accept the system for the task; and were they overall satisfied with use of the system for the task.

	Under control	Gaze correct	No delay	Mental demand	Eye tired	Accept-ance	Satisfact-ion
Selecting	4.57	4.64	3.43	2.21	2.64	4.00	4.00
Sliding	3.36	3.50	3.29	3.07	3.64	3.00	3.29
Scrolling	3.64	3.64	3.14	2.14	2.86	3.93	3.86

(a) Album cover browser.   (b) Gaze quiz.

Figure 10. Two SideWays applications used in the user study: *album cover browser* allows users to navigate music albums; *gaze quiz* presents a question in the center and uses the side inputs for "Yes/No" answers.

feedback data. Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Bonferroni correction applied.

We asked the participants whether they felt they were in direct control of the SideWays system. One participant particularly commented "I can feel a real power, starting from my brain and ending on the screen." We found a significant difference between the three tasks, $\chi^2(2)$=17.07, p<.001. A post-hoc analysis showed that the selecting vs sliding (p=0.002) and selecting vs scrolling (p=0.002) pairs were significantly different. The participants felt that they were most in control when using our system for selecting objects.

We found a significant difference in whether the participants perceived the system as responding correctly to their gaze, $\chi^2(2)$=14.70, p<.002. The selecting vs sliding (p=0.001) and selecting vs scrolling (p=0.012) pairs were significantly different. The participants felt that the system was most responsive when used for selecting objects.
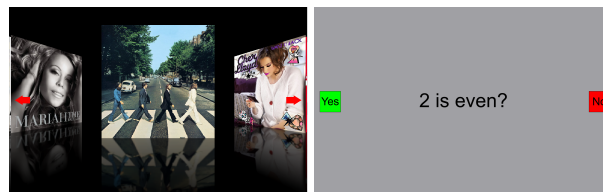
We also found a significant difference in tiredness of eyes in using *SideWays*, $\chi^2(2)$=7.54, p<.023. The participants felt significantly more tired when using our system for sliding than for selecting (p=0.006).

Participants provided comments in comparison of scrolling and sliding. Sliding was found demanding as it required target status check in the centre of the screen while concentrating gaze on either left or right. For scrolling, participants noted that they can see what is coming while they are controlling, and they found it less demanding to use peripheral vision to check what is in the centre, as it was displayed more largely than in the sliding task.

Participants also commented on possible applications of the system. Since SideWays is touch-free, the interaction is sanitary and therefore suitable for public environments, such as airports, libraries and shopping malls. Some participants suggested that it could benefit disabled people with paralysis. Several participants criticized the lack of visual feedback for the detection of gaze directions. This is important as it provides indication of whether the system interpreted the user's input correctly.

APPLICATIONS

After completing the previous three tasks, participants interacted with two SideWays applications (see Figure 10): an al-

bum cover browser and a gaze quiz. The album cover browser acts as an interface of a music jukebox. A user browses for music by scrolling left and right, and the centre region represents the music album to play. The gaze quiz application is an interactive quiz game. A user first reads a question displayed in the centre, and then answers the question by selecting yes or no, which was placed on the left and right positions of the screen, respectively. We used a window size 3 for the album browser and a window size 4 for the gaze quiz. Both applications used discrete activation. Only qualitative feedback of the system (e.g. preference and suggestions) was collected.

Participants were allowed to use the interfaces freely. We gave no instruction of how to interact with the interfaces, but the participants were still able to use the applications. For the media browser application, the users were able to navigate through all the music album covers and check what music was available. Sometimes, they scrolled back/forward to stop at the one they were interested in. During the gaze quiz application, the participants read twelve questions displayed in the centre and provided answers for all the questions. All participants understood how to search the media album covers and make selections. One participant encountered the "Midas Touch" problem in the gaze quiz [7]. While we displayed one question with a long sentence, the participant accidentally chose the answer "No" (on the right end) as he was reading the sentence. Thus, information in the centre should not be extended to the far left or right regions of the display.

Participants further suggested that SideWays could be applied in situations where a display is obstructed by a glass wall or window, such as shop displays for pedestrians. Another suggestion was for controlling television, e.g. adjusting volume or switching channels.

DISCUSSION

Our study validates that SideWays enables eye gaze as input for interactive displays, without the need of prior calibration or specialist hardware. This is significant in a number of ways. First, achieving robust gaze control, albeit coarse-grained, without need for calibration means that our system is person-independent. Any user can walk up to a display fitted with our system, and interact with it using their eyes only. Secondly, as we overcome calibration, users will be able to interact immediately (in principle) which is important for serendipitous and short-lived interactions that don't warrant preparation phases. Thirdly, we achieve gaze control with an

off-the-shelf camera. This means that displays can be made gaze-aware at low cost, potentially on pervasive scale.

Effects of System Parameters

The threshold T_r (used in equation 2) defines the size of the central region, and in our prototype was set to corresponds to a horizontal visual angle of $40°$. Most participants were aware how far they needed to look left/right as SideWays provided guidance by displaying the visual control stimulus. Increasing the threshold essentially increases the visual angle. If the threshold is small, the central region becomes narrower, and the system also becomes more sensitive to small eye movements of looking left/right. However, if the threshold is large, the user will need to look left/right further to trigger input which might cause discomfort and fatigue. The optimal threshold will depend on application, and designers need to consider the distance between the user and the display and the size of the display.

The window size determines how much evidence is collected before an action is triggered and trades off between accuracy and speed. A large window size improves the accuracy of gaze detection, but causes longer delays and slower response. This is suitable for discrete actions (e.g. selecting an object), where it is more important that the system detects the correct object. A small windows size speeds up response, but increases likelihood of errors caused by noise. This will be reasonable for continuous actions (e.g. sliding) where faster response is important and where the effect of occasional misclassifications will quickly corrected by continuous updates.

Designers can map input state to discrete versus continuous actions to fit the nature of the task. For example, if the content is visual (e.g. a photo album), continuous action may be chosen for fast scrolling as our study participants found that they can scroll to larger distinctive objects with peripheral vision. If the content requires attention (e.g., flicking through book pages), a discrete action mapping is better suited.

Limitations and Design Considerations

Responsiveness. When the system is not able to detect eyes images of sufficiently quality for computing eye centre and eye corners, the interface responsiveness decreases. This happens when users blink and squint or when the eyes get occluded in any other way, and can also be caused by larger head movement. When the system does not respond correctly or fast enough, no manual intervention is needed to reset SideWays. Participants reported that they reinitialized SideWays by closing their eyes, or by adjusting head positions or distance to the screen, indicating a good understanding of what causes misfunction and how to recover.

Head orientation. Our system requires users to keep their head oriented toward the centre of the display and only move their eyes. Some participants commented that this was unnatural, because they often subconsciously turn their head towards the direction of their visual focus. As a result, the detection of gaze direction becomes unreliable. This poses a limitation of user interaction. Although restricted head movement was commented as unnatural, in general, the participants were able to recover from loss of gaze detection by quickly correcting their head orientation, in order to achieve their tasks. To minimize head turning, designer should pay attention to the display region where information changes. Dynamic movement on the control regions can attract user attention; hence, causing head turning.

Provide detection feedback. Users need feedback when an event is triggered, to understand whether the system has detected their gaze. Visual feedback can be explicit confirmation of users input, for example by highlighting a displayed control that was triggered, or implicit in the behaviour of application, for example by updating the content displayed in the centre of the screen. However, when users glance sideways to trigger a control, it can be difficult for user to acquire feedback that effects only the centre of the display.

Toward use with Public Displays

Person-independence and interaction without preparation are critical steps toward genuinely spontaneous interaction with displays we encounter in public environment. While our evaluation shows that our system achieves both, it does so under the constraints of a controlled study designed to systematically test and characterise the interaction technique. Deployment in real-world contexts naturally raises a range of further challenges. For example, although a calibration phase is avoided, there will still be a gulf in how users can readily obtain and use gaze control over a display they encounter. However, with a lab study of our system's working, we now have a foundation for addressing deployment challenges, as well as insights on user performance and strategies that can inform application design.

For ecological validity, it will be important to inform further development by understanding of how users interact with displays "in the wild". For example, research on public display interaction showed that often more than one user is involved in interaction [12, 15], posing a challenge to accommodate multiple users simultaneously. For SideWays, the calculation of multiple users' gaze is equivalent to individual's gaze estimation in parallel. However, detection of eye images will naturally be more complex in public environments and with multiple users. Although the system would be able to discriminate input from different pairs of eyes, it would not be clear how to map multi-user input to display actions without causing potential conflict.

CONCLUSION

In this paper, we presented SideWays, a novel eye gaze interface for spontaneous interaction with a display. It leverages left and right regions of the display for gaze controls, while keeping the central region for display content. The system uses an off-the-shelf visible light camera along with its lightweight image processing for robust detection of gaze directions. We conducted a user study to evaluate SideWays on three interactive tasks. The results show that people are able to use SideWays to interact with a display, and we have gained insights of people's gaze control strategies. With SideWays's nature of being calibration-free and person-independent, its applications are potentially usable by many people. This

work represents a significant step toward use of our eyes for serendipitous interaction with pervasive displays.

REFERENCES

1. Brignull, H., and Rogers, Y. Enticing people to interact with large public displays in public spaces. In *Proc. INTERACT 2003*, IOS Press (2003), 17–24.

2. Duchowski, A. T. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag New York, Inc., 2003.

3. Eaddy, M., Blasko, G., Babcock, J., and Feiner, S. My own private kiosk: Privacy-preserving public displays. In *Proc. ISWC 2004*, IEEE Computer Society (2004), 132–135.

4. Geusebroek, J.-M., van den Boomgaard, R., Smeulders, A. W., and Geerts, H. Color invariance. *IEEE Trans. Pattern Anal. Mach. Intell. 23*, 12 (2001), 1338–1350.

5. Hansen, D., and Ji, Q. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell. 32*, 3 (2010), 478–500.

6. Hutchinson, T., White, K.P., J., Martin, W., Reichert, K., and Frey, L. Human-computer interaction using eye-gaze input. *Trans. Sys. Man Cyber Part C 19*, 6 (1989), 1527 –1534.

7. Jacob, R. J. K. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans. Inf. Syst. 9*, 2 (1991), 152–169.

8. Kumar, M., and Winograd, T. Gaze-enhanced scrolling techniques. In *Proc. UIST 2007*, ACM Press (2007), 213–216.

9. Magee, J. J., Betke, M., Gips, J., Scott, M. R., and Waber, B. N. A human-computer interface using symmetry between eyes to detect gaze direction. *Trans. Sys. Man Cyber. Part A 38*, 6 (2008), 1248–1261.

10. Morimoto, C. H., and Mimica, M. R. M. Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst. 98*, 1 (2005), 4–24.

11. Mubin, O., Lashina, T., and Loenen, E. How not to become a buffoon in front of a shop window: A solution allowing natural head movement for interaction with a public display. In *Proc. INTERACT 2009*, Springer-Verlag (2009), 250–263.

12. Müller, J., Walter, R., Bailly, G., Nischt, M., and Alt, F. Looking glass: a field study on noticing interactivity of a shop window. In *Proc. CHI 2012*, ACM Press (2012), 297–306.

13. Nakanishi, Y., Fujii, T., Kiatjima, K., Sato, Y., and Koike, H. Vision-based face tracking system for large displays. In *Proc. UbiComp 2002*, Springer-Verlag (2002), 152–159.

14. Noris, B., Benmachiche, K., and Billard, A. Calibration-free eye gaze direction detection with gaussian processes. In *Proc. VISAPP 2008*, INSTICC (2008), 611–616.

15. Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., and Saarikko, P. It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In *Proc. CHI 2008*, ACM Press (2008), 1285–1294.

16. San Agustin, J., Hansen, J. P., and Tall, M. Gaze-based interaction with public displays using off-the-shelf components. In *Ubicomp 2010 Adjunct*, ACM Press (2010), 377–378.

17. Sesma, L., Villanueva, A., and Cabeza, R. Evaluation of pupil center-eye corner vector for gaze estimation using a web cam. In *Proc. ETRA 2012*, ACM Press (2012), 217–220.

18. Sibert, L. E., and Jacob, R. J. K. Evaluation of eye gaze interaction. In *Proc. CHI 2000*, ACM Press (2000), 281–288.

19. Sippl, A., Holzmann, C., Zachhuber, D., and Ferscha, A. Real-time gaze tracking for public displays. In *Proc. AmI 2010*, Springer-Verlag (2010), 167–176.

20. Smith, J. D., Vertegaal, R., and Sohn, C. Viewpointer: lightweight calibration-free eye tracking for ubiquitous handsfree deixis. In *Proc. UIST 2005*, ACM Press (2005), 53–61.

21. Sugano, Y., Matsushita, Y., and Sato, Y. Calibration-free gaze sensing using saliency maps. In *Proc. CVPR 2010*, IEEE Computer Society (2010), 2667–2674.

22. Valenti, R., and Gevers, T. Accurate eye center location and tracking using isophote curvature. In *Proc. CVPR 2008*, IEEE Computer Society (2008), 1–8.

23. Vertegaal, R., Mamuji, A., Sohn, C., and Cheng, D. Media eyepliances: using eye tracking for remote control focus selection of appliances. In *Ext. Abstracts CHI 2005*, ACM Press (2005), 1861–1864.

24. Vertegaal, R., Shell, J. S., Chen, D., and Mamuji, A. Designing for augmented attention: Towards a framework for attentive user interfaces. *Computers in Human Behavior 22*, 4 (2006), 771 – 789.

25. Viola, P., and Jones, M. Rapid object detection using a boosted cascade of simple features. In *Proc. CVPR 2001*, IEEE Computer Society (2001), 511–518.

26. Ward, D. J., and MacKay, D. J. C. Fast hands-free writing by gaze direction. *Nature 418*, 6900 (2002), 838.

27. Zhang, Y., Bulling, A., and Gellersen, H. Towards pervasive eye tracking using low-level image features. In *Proc. ETRA 2012*, ACM Press (2012), 261–264.

28. Zhu, Z., and Ji, Q. Eye and gaze tracking for interactive graphic display. *Mach. Vision Appl. 15*, 3 (2004), 139–148.

29. Zhu, Z., and Ji, Q. Eye gaze tracking under natural head movements. In *Proc. CVPR 2005*, IEEE Computer Society (2005), 918–923.